

Reg. No.:

Name :



Mid Term Examination – March 2015

Programme	: M.Tech Computer Science & Engineering	Semester	: Winter 2015
Course	: Natural Language Processing	Code	: CSE528
Faculty	: Prof. Tulasi Prasad Sariki	Slot	: F2
Time	: One and Half Hours	Max. Marks	: 50

Answer ALL Questions

Q.No.	Sub. Sec.	Question Description	Marks
1.		An organization is trying to extract all their customer email addresses from a text document. Help them out by writing a regular expression to fulfill their need.	6
2.		Deduce the reasons for using Finite State transducer rather than the finite state automaton for developing morphological analyzers?	4
3.		Choose the correct POS tag out of choices for each word in the following sentences a. She[PRP] promised[VBD, VBN] to[IN, TO] back[NN, RB, VB, JJ] the[DT] bill[NN, VB]. b. Fruit[NN, VB] flies[NNS, VBZ] are[VB] attracted[VBD, VBN] by[IN, RP] ripened[VBD, VBN] fruits[NNS] and[CC] vegetables[NNS]. c. He[PRP] picked[VBD, VBN] up[JJ, RP, VB, NN, IN] the[DT] block[NN, VB] of[In] wood[JJ, NN, VB] and[CC] began[VBD] to[IN, TO, RP] work[NN, VB, JJ] on[IN, RP, JJ] it[PRP]. d. Could[MD] you[PRP] hand[RB, NN, VB] me[PRP\$] that[CC, PRP, RB, WDT] bit[NN, VB] for[CC, IN] this[RB, PRP, JJ] drill[VB, NN]? e. If[NN, CC] you[PRP] look[VB, NN] up[IN, VB, NN, RP, JJ] there[RB, EX, JJ] you[PRP] can[NN, MD] see[NN, VB] a[DT] bat[NN, VB] flying[NN, VBG] between[IN, RP] the[DT] trees[NNS, VBZ].	10

4.	State any five challenges of sentiment analysis and suggest your mechanism to overcome those challenges.	10
5.	You are given a large text document to produce effective summary. Describe your methodology to summarize in more effective way.	10
6.	An organization need a stemmer to perform some text analytics, suggest a suitable approach from the given choices i. Table look-up approach ii. Successor Variety iii. n-gram iv. Affix Removal State the reasons behind your choice.	5
7.	Suppose you want to carry out corpora analysis of English for identifying the distribution of lengths of sentences in the corpus but the sentence boundaries are not annotated in the corpus. Explain any one approach to carry out this task?	5
