Reg. No.:

Name   :

Mid Term Examination KEY − 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. *{^[A−Za−z0−9._ −]+@[[A−Za−z0−9.−]+$}*	6
2.		Deduce the reasons for using Finite State transducer rather than the finite state automaton for developing morphological analyzers? *FSA can act as a* 　*Recognizer* 　*Generator* 　*5 tuple Representation* 　*Equivalent to regular languages*　　　*FST can act as a* 　　　　　　　　　　　　　　　　　　　*Recognizer* 　　　　　　　　　　　　　　　　　　　*Generator* 　　　　　　　　　　　　　　　　　　　*7 tuple Representation* 　　　　　　　　　　　　　　　　　　　*Equivalent to regular relations* 　　　　　　　　　　　　　　　　　　　*Translator* 　　　　　　　　　　　　　　　　　　　*Set relator*	4

| 3. | Choose the correct POS tag out of choices for each word in the following sentences | 10 |
| | a. She[PRP] promised[VBD,VBN] to[IN,TO] back[NN,RB,VB,JJ] the[DT] bill[NN,VB]. | |

a. She[PRP] promised[VBD,VBN] to[IN,TO] back[NN,RB,VB,JJ] the[DT] bill[NN,VB].

She/PRP promised/VBD to/TO back/VB the/DT bill/NN

b. Fruit[NN,VB] flies[NNS,VBZ] are[VB] attracted[VBD,VBN] by[IN,RP] ripened[VBD,VBN] fruits[NNS] and[CC] vegetables[NNS].

Fruit/NN flies/NNS are/VB attracted/VBN by/IN ripened/VBD 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].

He/PRP picked/VBD up/RP the/DT block/NN of/IN wood/NN and/CC began/VBD to/TO work/VB on/IN 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]

Could/MD you/PRP hand/VB me/PRP that/WDT bit/NN for/IN this/PRP drill/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].

If/CC you/PRP look/VB up/RP there/RB you/PRP can/MD see/VB a/DT bat/NN flying/VBG between/IN the/DT trees/NNS

| 4. | State any five challenges of sentiment analysis and suggest your mechanisms to overcome those challenges. | 10 |

i. *Named Entity Recognition – What is the person actually talking about, e.g. is 300 Spartans a group of Greeks or a movie?*

ii. *Anaphora Resolution – the problem of resolving what a pronoun, or a noun phrase refers to. "We watched the movie and went to dinner; it was awful." What does "It" refer to?*

iii. *Parsing – What is the subject and object of the sentence, which one does the verb and/or adjective actually refer to?*

iv. *Sarcasm – If you don't know the author you have no idea whether 'bad' means bad or good.*

v. *Twitter – abbreviations, lack of capitals, poor spelling, poor punctuation, poor grammar.*

vi. *Fake reviews/Spam*

vii. *Biased reviews*

viii. *Word sense disambiguation*

ix. *Preprocessing and Cleaning*

5.	You are given a large text document to produce effective summary. Describe your methodology to summarize in more effective way.
	
	    i. *Statistical*
	    ii. *Linguistics (NLP)*
	    iii. *Lexical chains*
		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.
	   Describe your choice and state the advantages of your choice over other
		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? *Sentence boundary disambiguation*	5

*****