M.C.S.E. Examination, 2017
(2nd Semester)
NATURAL LANGUAGE PROCESSING

Time: Three Hours **Full Marks : 100**

Answer any *five* questions.

1. (a) Define the string patterns that are identified by the regular expression /(^|[^a-zA-Z])[tT]he[^a-zA-Z]/ 3
   (b) Write an Finite State Automata for time-of-day expressions like *eleven o'clock, twelve noon, twelve midnight, twelve-thirty, thirty minutes past twelve, a quarter to ten, a quarter past ten* etc. 7
   (c) Morphological parsing can be more complicated than morphological generation because of the problem of ambiguity. Give an example. 2
   (d) Why morphotactics and orthographic rules are required in a morphological parser? 2 + 2
   (e) Explain the following Porter Stemmer rules (*v*) Y → I and (m>0) ALITI → AL. Now, show that the stemmer will generate 'formal' from the word 'formality'. 4

2. (a) What is the noisy channel approach for correcting non-word spelling errors in terms of likelihood and prior probability? 4
   (b) Define the likelihood probability in terms of the following four confusion matrices: 8
   del[x,y]: count(xy typed as x)
   ins[x,y]: count(x typed as xy)
   sub[x,y]: count(x typed as y)
   trans[x,y]: count(xy typed as yx)
   (c) How real word spelling errors can be solved by noisy channel model? 2
   (d)Each cell D(i,j) in the Edit-distance matrix contains the distance between the first i characters of the target string and the first j characters of the source string. String transformation can be caused by insertion, deletion or substitution operations. Write the recurrence relation for computation of D(i,j) in terms of the insertion, deletion and substitution costs. 6

3. (a) In discriminative model for classification the weight-feature dot product produces values ranging from $-\infty$ to $+\infty$. In order to restrict the weight-feature dot product value to range between 0 to 1, suggest techniques and hence develop the equation for computing $p(c|x)$ in logistic regression model. 6
Show that the best assigned class to x is the one that produces highest dot product between the weights and the features. 2

(b) How the tag transition and word emission probabilities be calculated in a trigram model of the HMM based POS tagger? 6

(c) How a POS tagger can be analyzed for error? 4

(d) How do you handle unknown words in a POS tagger. 2

4. (a) How do you define perplexity of a language model? 4

(b) How a language model handles unknown words in an open vocabulary system? 2

(c) What is Stupid backoff in web-scale language model? 2

(d) The equation for interpolated absolute discounting applied to bigrams includes two terms, the first one is the discounted bigram while the second one is the unigram with the interpolation weight λ. Discuss. 4

(e) In Kneser-Ney smoothing, instead of the unigram probability, the $P_{CONTINUATION}(W)$ probability is considered which is defined as 'how likely W to appear as a novel continuation'. How do you calculate $P_{CONTINUATION}(W)$? 4

(f) Define and discuss the final equation for interpolated Kneser-Ney smoothing. 4

5. (a) Consider the following English Context Free Grammar:

S -> NP VP	NP -> Pronoun
S -> Aux NP VP	NP -> Possessive_Pronoun Noun
S -> VP	NP -> NP PP
NP -> Det Nominal	PP -> Preposition NP
Nominal -> Noun	VP -> Verb
Nominal -> Noun Nominal	VP -> Verb NP
NP -> Det Noun	VP -> Verb NP PP

What is the left-corner table that can be used for bottom-up filtering in a top-down parser for the above grammar? What is the role of the left corner table? Identify the production rule that causes left recursion. 8

(b) The CKY parsing algorithm requires that the grammar should be in Chomsky Normal Form (CNF). What are the forms of the grammar rules in CNF? How can the following grammar rule be converted to CNF: S → Aux NP VP where Aux, NP and VP refer to the auxiliary, noun phrase and verb phrase respectively. 6

(c) Consider the following Gold standard brackets for the sentence 'Sales executives were examining the figures with great care yesterday' :
S(0,11), NP(0,2), VP(2,9), VP(3,9), NP(4,6), PP(6,9), NP(7,9), NP(9,10).
The candidate brackets as output of the Probabilistic Context Free Grammar (PCFG) Parser for the same sentence are :
S(0,11), NP(0,2), VP(2,10), VP(3,10), NP(4,6), PP(6,10), NP(7,10).
Now, define and calculate the Labeled Precision, Labeled Recall and LP/LR F1 Score for the above PCFG parser. 6

6. (a) When a dependency tree is said to be projective? Why Transition based Dependency parsing produces errors while parsing sentences with non-projective structures? 2+2

(b) Consider the sentence 'Book me the flight through London'.
The reference dependency parse for the sentence is as follows:
root(Book), iobj(Book, flight), obj(Book, me), det(flight,the), nmod(flight, London) and case(London, through).
The system generated dependency parse for the sentence is as follows:
root(Book), x-comp(Book, flight), nsubj(flight,me), det(flight, the), nmod(flight, London) and case(London, through).
Now, define Labeled Attachment Score (LAS) and Unlabeled Attachment Score (UAS) evaluation metrics and calculate the same for the above scenario. 6

(c) A configuration in transition based parsing consists of a stack, an input buffer of words or tokens and a set of relations representing a dependency tree.
Define the three transition operators LEFTARC, RIGHTARC and SHIFT that operate on the top two elements of the stack. 3
Consider the sentence 'Book me the flight through London'. Show the transition steps in generating the dependency tree using transition based parsing. Note that the reference dependency parse for the sentence is given in (b). 7

7.(a) How do you define Pointwise Mutual Information (PMI) between a word w and its context c? What is the rationale for defining the Positive PMI?
PMI has the problem of being biased towards infrequent events. Suggest a solution for overcoming the problem. 2+2+2

(b) Discuss how Singular Value Decomposition is applied to word-word or word-context matrices. 6

(c) Describe the skip-gram model for learning word and context embeddings in terms of the input layer, projection layer and output layer. 8

8.(a) Define the synonymy, anotonymy, hypernymy and meronymy relations among word senses with examples. 4

(b) Describe the Resnik method for computing the similarity between two words using information content. 6

(c) What is the extended Lesk algorithm for computing the similarity between two words? 4

(d) Describe the collocational and bag of words features in supervised word sense disambiguation (WSD) task. Discuss how Naïve Bayes method can be applied to WSD task. 2+4

---------- x ----------