Total No. of printed pages = 4

CS 1317 E 012

Roll No. of candidate

2018

B.Tech. 7th Semester End-Term Examination

NATURAL LANGUAGE PROCESSING

(Elective-I)

Full Marks – 100   Time – Three hours

The figures in the margin indicate full marks
for the questions.

Answer Q.No. 1 and any *six* from the rest.

1.   Answer the following questions :   (10 × 1 = 10)

   (i)   ————————— is the study of the way words are built up from smaller meaning bearing units, morphemes.

   (ii) ————————— is the term used for automatic recognition of machine or hand-printed characters.

   (iii) ————————— are the class of probabilistic models that assume that we can predict the probability of some future unit without looking too far into the past.

   (iv) The task or reevaluating some of the zero-probability and low-probability N-grams and assigning them non-zero values is called ————————.

(v) The task of ————— is to examine word tokens in context and specify exactly which sense of word is being used.

(vi) ————— is a corpus that is annotated with verbal propositions and their arguments.

(vii) In machine learning, ————— approaches use collections of texts annotated their correct senses to train classifiers.

(viii) The ————— views documents and queries as vectors in a large multidimensional space.

(ix) There are two broad classes of ways to form words from morphemes: ————— and derivation.

(x) A regular expression is one way of characterizing a particular kind of formal language called —————.

2. (a) Briefly explain the Unicode system for text representation in computers. (5)

(b) How Transformation Based Tagging rules are applied? Give example. (3 + 2 = 5)

(c) Explain the issues in computational morphology with suitable example. (4 + 1 = 5)

3. (a) What do you mean by Formal language? Give example. (3 + 2 = 5)

(b) Explain lexemes and phonemes in natural language processing. (2.5 × 2 = 5)

(c) Explain the different regular expression patterns. (5)

4. (a) What is Add One Smoothing? Explain with example. (4)

(b) Explain the importance of n-gram model in word prediction. Give example. (5)

(c) Define backoff and interpolation in n-gram model. (3 + 3 = 6)

5. (a) What do you mean by unknown words in a POS tagging task? How are POS tags assigned to unknown words? (4 + 4 = 8)

(b) Briefly explain the Porter Stemmer. (4)

(c) Describe the various tuples of Finite State Transducer. (3)

6. (a) Briefly explain bigram model for spelling correction. (4)

(b) Explain linear regression model. Give example. (3 + 2 = 5)

(c) What is top down parsing? What is ambiguity in parsing? (3 + 3 = 6)

7. (a) What do you mean by Homonymy, Polysemy and Synonymy for information retrieval systems? (2 × 3 = 6)

(b) Explain Vector Space model of information retrieval. Give example. (3 + 2 = 5)

(c) Explain the feature vector for machine learning approaches in word sense disambiguation. (4)

8. (a) Explain the different phases in natural language processing. (2 × 5 = 10)

(b) Explain the regular expression: /[^a-zA-Z][tT]he/ (3)

(c) Differentiate between NFA and DFA. (2)

9. (a) Explain the different types of Morphology. Give examples. (2 × 4 = 8)

(b) What are the different categories of single word misspellings? (4)

(c) Mention the different elements in balanced corpus. (3)

———————