UNIT-V

9. (a) Given below the four documents (each document is just one sentence) 8

    1. Kitty is a cat.
    2. Kitty sat and ate a rat.
    3. Rats eat cheese.
    4. A rat is sitting

   (i) What will be the content of each document after "stop" word removal?

   (ii) What will be the content of each document after "stemming" each word in the document after stop word removal?

   (iii) What is the vocabulary after stop word removal and stemming?

   (iv) Create a tf-idf vector for each document. Do this after stop word removal and stemming.

   (b) What are the three classic models in information retrieval system? 2

   (c) What are the benefits of eliminating stop words? Give examples in which stop word elimination may be harmful 4

10. Write short notes on 5+4+5=14

   (i) Application of NLP

   (ii) Text Summarization

   (iii) Vector Space Model

****

M. Tech Odd Semester Examination, February, 2023

Computer Science & Engineering

(1st Semester)

Course No.: MCSE-104(B)

(Natural Language Processing)

Full Marks: 70

Pass Marks: 28

Time: 3 hours

Note: 1. Attempt 05 (Five) questions by taking one form each unit.
2. Begin each answer in a new page.
3. Answer parts of a question at a place.
4. Assume reasonable data wherever required.
5. The figures in the right margin indicate full marks for the question.

UNIT–I

1. (a) What is Natural Language Processing? What are the challenges of Natural Language Processing 5

   (b) Explain with an example different levels of Natural Language Processing. 9

2. (a) Write the differences betweeni) Finite state automata and Finite state Transducer 3

   (b) Briefly, explain the morphology operations: compounding, derivation and inflection. Given the root 'snow', give an example each of the result of applying the three morphology operations 6

(c) Consider the regular expression below which is supposed to describe some morphological facts about English adjectives

(un-)*(adj-root)(-er | -est | -ly)*

What is the problem with the proposed regular expression model? And how will you go about correcting it - explain your approach. 5

UNIT-II

3. (a) What is language modeling? What are the approaches to language modeling. 2+3=5

(b) What are the metrics used to evaluate N-grams systems? 3

(c) What is smoothing? Explain Laplace smoothing technique. 2+4=6

4. (a) Write the CYK Algorithm. Given the grammar below in Chomsky normal form. Fill up a CYK chart for the sentence : The tall boy hit the ball.

Grammar (start symbol is S): S→NP VP, VP → V NP, NP → DET N, N → ADJ N, V → hit, DET → the, N → boy | ball, ADJ→ tall. 4+4=8

(b) What is verb subcategorization?. 2

(c) Identify the morphological type (Adverb phrase, Verb Phrase,Adjective Phrase, Preposition Phrase) of following sentence segments

1. rapidly like a bat
2. play the piano
3. hungry as a horse
4. across the line 4

UNIT-III

5. (a) Explain Hidden Markov Model tagger for parts of speech tagging. 5

(b) What is over-generation problem? How can we solve the problem of over-generation using feature structure and unification? 1+4=5

(c) What are different types of sentence structure? 4

6. (a) What are elements of First Order Predicate Logic? Give examples 5

(b) What is WordNet? Explain with examples the relations in WordNet. 1+8=9

UNIT-IV

7. (a) What is reference? What are the different types of references? Explain them with examples 10

(b) Differentiate between cohesion and coherence. 2

(c) What are lexical cohesion devices? 2

8. (a) Explain with example a pronoun resolution algorithm 10

(b) What is word sense disambiguation? What are the different approaches of word sense disambiguation. 4