# MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

*A Constituent Institution of Manipal University*

**VII SEMESTER B.TECH.**

**END SEMESTER EXAMINATIONS, NOVEMBER/DECEMEBER 2017**

SUBJECT: NATURAL LANGUAG PROCESSING (CSE 4011)

**REVISED CREDIT SYSTEM**

**(25/11/2017)**

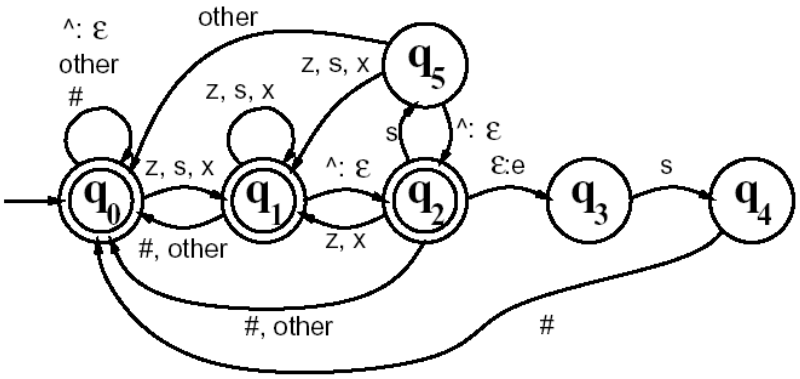
Time: 3 Hours MAX. MARKS: 50

**Instructions to Candidates:**

- Answer **ALL** the questions.
- Missing data may be suitably assumed.

1A. Draw a Finite State Automaton (FSA) which can accept only the following words. The FSA must contain minimal number of states. Show the numbering of states in your FSA. *cmos, capacitors, resistors, respires, transistors, transpires* 4

1B. The E-insertion rule states that an *e* is to be added after –s, -z, -x , -ch , –sh before –s. The following Transducer handles only –s, -z, -x cases. Extend the FST to handle –ch, –sh cases.




3

1C. Tabulate the knowledge that is required to process written text and spoken text separately. 3

2A. Enumerate the steps involved in Rule based tagger and Transformational tagger by taking "Time" as an example word in a sentence "Time flies like an arrow". State one similarity and one difference between the Rule based Tagger and the Transformational Tagger. 4

2B. What kinds of linguistic phenomena can be captured in bigrams? For a word *w* with its count *c*, write expressions to compute the following probabilities:

i) Maximum Likelihood estimate

ii) Laplace estimate

iii) Adjusted count

iv) Relative discount 3

2C. Assume a teacher has asked students to write a play in the style of Shakespeare. It is needed to score their plays using a trigram language model computed from a corpus of all Shakespeare plays. But the data is too sparse and most of students' sentences receive a score of zero. How would you use a back-off model to alleviate this problem? How about Interpolation? Also, state the necessary formulations. 3

3A. Derive an expression to compute the most probable tag sequence for the sequence of words using bigram HMM tagging. State all assumptions clearly. 4

3B. What is the tagging of the following sentence, "computers process programs accurately" with the following HMM tagger:
computers N 0.123
process N 0.1
process V 0.2
programs N 0.11
programs V 0.15
accurately Adv 0.789
P(N|V)=0.5 P(N|Adv)=0.12 P(V|Adv)=0.05
P(V|N)=0.4 P(Adv|N)=0.01 P(Adv|V)=0.13
P(N|N)=0.6 P(V|V)=0.05 3

3C. What is the problem addressed by a Part-of-Speech (PoS) tagger? Why is it not trivial? What are the two main difficulties? 3

4A. Construct CKY parsing table for the given grammar and the sentence "the frogs ate fish". State the rules that are needed to fill out cell [i, j]

| | |
|---|---|
| S → NP VP | Det → a \| the |
| NP → Det Nom | N → fish \| frogs \| soup |
| NP → Nom | Prep → in \| for |
| Nom → N SRel | TV → saw \| ate |
| Nom → N | IV → fish \| swim |
| VP → TV NP | Relpro → that |
| VP → IV PP | |
| VP → IV | |
| PP → Prep NP | |
| SRel → Relpro VP | |

5

4B. Compute the most probable parse tree for the sentence "Can you book Boeing flights" using the PCFG given below. Clearly show all the steps.

| | |
|---|---|
| S → NP VP [.80] | Det → that [.05] \| the [.80] \| a [.15] |
| S → Aux NP VP [.15] | Noun → book [.10] |
| S → VP [.05] | Noun → flights [.50] |
| NP → Det Nom [.20] | Noun → meal [.40] |
| NP → Proper-Noun [.35] | Verb → book [.30] |
| NP → Nom [.05] | Verb → include [.30] |
| NP → Pronoun [.40] | Verb → want [.40] |
| Nom → Noun [.75] | Aux → can [.40] |
| Nom → Noun Nom [.20] | Aux → does [.30] |
| Nom → Proper-Noun Nom [.05] | Aux → do [.30] |
| VP → Verb [.55] | Proper-Noun → Boeing [.40] |
| VP → Verb NP [.40] | Proper-Noun → Delhi [.40] |
| VP → Verb NP NP [.05] | Pronoun → you [.40] \| I [.60] |

3

4C. Distinguish between Recognizer and Parser with respect to CKY Parsing table. 2

5A. Draw a parse tree in the inverted form representing a parse for the given sentence. Clearly show all the edges corresponding to each word and arrive at the start symbol S from the given sentence. Make a list of production rules as per English grammar. Consider the following tag set for words. State the Grammar rule for noun phrase denoted as NP in terms of pre determiners and post determiners.
Noun – NN, Verb – VBD, preposition – IN, Adjective – JJ
*Economic news had little effect on financial markets* 4

5B. Describe the importance of statistical MT approaches and formulate the expression to model the translation in terms of Bayes Rule. 3

5C. Discuss the ambiguities faced in Named Entity Recognition systems with an example for each. 3