Question Bank

Sub : Natural Language processing (**DLO8012)**

****

Module 1: Introduction

Q.1 What is Natural language processing (NLP) ? Discuss various stages involved in NLP process with suitable example.

OR

What is Natural Language Understanding? Discuss various levels of analysis under it with example.

Q.2 What do you mean by ambiguity in Natural language? Explain with suitable example. Discuss various ways to resolve ambiguity in NL.

OR

What do mean by lexical ambiguity and syntactic ambiguity in Natural language? What are different ways to resolve these ambiguities?

Q.4 Discuss various challenges in processing natural language.

Q.5 List various applications of NLP and discuss any 2 applications in detail.

Module 2: Word-Level Analysis

Q.1 what is morphology. Why do we need to do Morphological Analysis? Discuss various application domains of Morphological Analysis.

Q.2 Explain derivational & inflectional morphology in detail with suitable examples.

OR

What are morphemes? What are different ways to create words from morphemes?

Q.3 what do you mean by stemming. Explain Porter's stemming algorithm in detail.

Q.4 What is the need of preprocessing for text data in natural language? Explain the steps of preprocessing (Tokenization , stopword removal and stemming) with example.

Q.5 what is language model? Explain the use of Language model?

OR

Write a note on N-Gram language Model.

Q.6 Problem on Ngram model

Consider following Training data:

<s> I am Sam </s>

<s> Sam I am </s>

<s> Sam I like </s>

<s> Sam I do like </s>

<s> do I like Sam </s>

Assume that we use a bigram language model based on the above training data.

What is the most probable next word predicted by the model for the following word sequences?

(1) <s> Sam ...

(2) <s> Sam I do ...

(3) <s> Sam I am Sam ...

(4) <s> do I like ...

(**Solution: Bigram probabilities: P(Sam|<s>) = 3/5 P(I|<s>) = 1/5 P(I|Sam) = 3/5 P(</s>|Sam) = 2/5 P(Sam|am) = 1/2 P(</s>|am) = 1/2 P(am|I) = 2/5 P(like|I) = 2/5 P(do|I) = 1/5 P(Sam|like) = 1/3 P(</s>|like) = 2/3 P(like|do) = 1/2 P(I|do) = 1/2**

(1) **and (3): "I". (2): "I" and "like" are equally probable. (4): </s>**)

Q.7 what is the role of FSA in Morphological analysis? Explain FST in detail.

Q.8 Design a Finite State Automata(FSA) for the words of English numbers 1-99.

Q.9 What do you mean by stemming? Explain Porter's stemming algorithm in detail and apply on following words to get the stems:

(i) Tapping

(ii) Smiling

(iii) Computerization

Module 3: Syntax Analysis

Q.1 what is POS tagging? Explain types of word classes in English NL. Also comment on possible tag sets available in ENGLISH NL

.OR

Explain open and closed word classes in English Language. Comment on possible tag sets available in ENGLISH NL. Show how the tags are assigned to the words of following sentence:

“Time flies like an arrow.”

Q.2 Why POS tagging is hard? Discuss possible challenges faced while performing POS tagging.

Q.3 Discuss various approaches to perform POS tagging.

Q.4 Explain in detail Rule based POS tagging/ Stochastic (HMM) POS tagging/ Hybrid POS tagging.

Q.5 Explain use of CFG in Natural Language Processing with suitable example.

Q.6 What is parsing? Explain/Compare Top-down & Bottom-up approach of parsing with suitable example.

Q.7 W.r.t following Grammar show Shift reduce parsing of following sentences

Book that flight

Does that flight include meal

S → NP VP S → Aux NP VP S → VP NP → Det NOM

NOM → Noun NOM → Noun NOM VP → Verb

VP → Verb NP

Det → that | this | a | the Noun → book | flight | meal | man

Verb → book | include | read Aux → does

Q.8 Construct a parse tree for the following sentence using CFG rules: “The man read this book”

Q.9 Explain how HMM/MaxEn/CRF is used for sequence labelling.