NLP QUESTION BANK

Module-1

1.What is Natural language processing (NLP)?

2.Discuss various stages involved in NLP process with suitable example.

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

with example. [Levels/ Stages-same]

4.What do you mean by ambiguity in Natural language? Explain with suitable example.

5.Discuss various ways to resolve ambiguity in NLP

6.What do mean by lexical ambiguity and syntactic ambiguity in Natural language? What

are different ways to resolve these ambiguities?

7.Discuss various challenges in processing natural language.

8.List various applications of NLP and discuss any 2 applications in detail. Explain pre-processing operation/steps in NLP: Tokenization, Stop word removal,script validation, filtration

9.Explain block diagram of generic NLP system

Module-2

1.What is Morphology?

2.Why do we need to do Morphological Analysis

3.Discuss various application domains of Morphological Analysis.

4.Define Morphemes and give examples.

5.What is mean by Bound & Free Morphemes Define Affixes and its type.

6.Difference between Orthographic & Morphological Rules.

7.What is Morphotactics

8.Define Inflectional and Derivational Morphology

9.What are different ways to create words from morphemes

10.What is Morphological Analysis? What do you mean by stemming.

11.Explain Porter's stemming algorithun in detail Apply on following words to get the stems:

1.Tapping

2.Smiling

3.Computerization

14.What is language model? Write a note on evaluation metrics for language model

15. Explain the use of Language model/ Applications of LM?

16.Write a note on N-Gram language Model.

17. Exercise on Bigram, Trigram, four gram exercises

18.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

19.What is the role of FSA in Morphological analysis? Explain FST in detail.

20. For a corpus, MLE for bigram “battery life” is 0.27, frequency of “battery is 800. After applying Laplace smoothening the MLE for “battery life” becomes 0.025. What is the vocabulary size of the corpus?

21.Design a Finite State Automata(FSA) for the words of English numbers 1-99. Write FSA for noun, verb and adjective.

22.Define Regular Relation

23.Write FST for regular and plural noun.

24.Define FST

25.FSA exercises related to NLP problems

26.Exercise on steming by porter stemmer, n-gram, k-gram, laplace smoothing good turing, FSA, FST

27. Find the probability of the following sentences

1.<s> Michael and Zack played at the playground </s>

2.<s> Bob went to the school </s>

3.<s> The school was huge</s>

4.<s>Zack went to the playground</s>

from the following corpus. Assume a bigram Language model(without Laplace smoothening).

<s> the school was open </s>

<s> Michael and Zack went to the school </s>

<s> the playground at the school was huge </s>

<s> Bob and Zack played at the playground </s>

<s> Bob Michael and Zack were friends </s>

Also find the Perplexity of the sentence which has highest probability.

28. Define Minimum edit distance used in Spelling correction.