A.Y 19-20 Even SEM

NLP Question Bank: End Semester Examination

Class: BE COMP (A+B)

1 Mark Sample Questions

1. Google Translate is one of the _____ applications.
a. Machine translation
b. Information Retrieval
c. Information Extraction
d. Summarization

2. Study of the meaning of words that are associated with grammatical structure is known as ----
   a. Morphology
   b. Phonology
   c. Semantics
   d. Reasoning

3. Consider the sentence "Horse ran up the hill. It was very steep. It soon got tired ". Identify the type of ambiguity.
   a. Syntactic ambiguity
   b. Lexical ambiguity
   c. Discourse ambiguity
   d. Pragmatic ambiguity

4. Identify the type of ambiguity in given sentence. " I Love you too "
   a. Syntactic ambiguity
   b. Lexical ambiguity
   c. Discourse ambiguity
   d. Pragmatic ambiguity

5. -------- it is also known as the parsing
   a. Lexical analysis

b. Syntactic analysis
c. Semantic analysis
d. Pragmatic analysis

6. ------ attempt to perform semantic and pragmatic processing of spoken utterance to understand what user is saying and what is being said.
   a. Language Translator
   b. Speech Recognition
   c. Information retrieval
   d. Information extraction

7. Which one is not the approach of Machine Translation
   e. a. Direct Translation
   f. b.Statistical based
   g. c. Image based
   h. d. Example based

8. Which of following is not use case to be solved using NLP techniques?
   a. Grammar or Sentence correction
   b. Image captioning
   c. Document Ranking
   d. Disease Prediction

9. The words 'there' and 'their' causes which of the following type of ambiguity?
   a. Syntactic
   b. Semantic
   c. Phonological
   d. Pragmatic

10. What is full form of NLG?

    a. Natural Language Generation
    b. Natural Language Genes
    c. Natural Language Growth
    d. Natural Language Generator

11. 28. Identify the ambiguity of statement "By books for children "

a. Lexical ambiguity

b. Syntactic ambiguity

c. Semantic ambiguity

d. Discourse ambiguity

12. The process of removing words like "and", "is", "a", "an", "the" from a sentence is called as
    a. Stemming
    b. Lemmatization
    c. Stop word Removal
    d. POS
13. The process of converting a sentence or paragraph into tokens is referred----
    a. Stemming
    b. Lemmatization
    c. Tokenization
    d. POS
14. Identify the odd one out from following
    a. nltk
    b. scikit learn
    c. SpaCy
    d. BERT
15. Which one of the following is not a pre-processing technique in NLP (R)
    a. Stemming and Lemmatization
    c. removing punctuations
    d. removal of stop words
    e. Sentiment analysis
16. Which of the following is an advantage of Porter stemmer over a full morphological parser?

a. The stemmer is better justified from a theoretical point of view

b. The output of a stemmer is always a valid word

c. The stemmer does not require a detailed lexicon to implement

d. It is most aggressive algorithm

17. Identify incorrect statement with respect to regular expressions (RE) in NLP

a. It is used for pattern matching

b. It is theoretical computer science technique

c. It is used for Linguistic

d. It is not supported in programming Language

18. Finite automata is called NFA when there is ------for specific input from current state to next state. (R)

a. single path

b. Multiple paths

c. No path

d. Only two paths

19. Which algorithm is not unsupervised algorithm?

a.Kmeans

b. k-Medoids

c.SVM

d. DBSCAN

20. Finite state transducer is ---- automation which recognizes or generate the pair of strings

      a. One tape

      b. Two tape

      c. Zero tapes

      d. Infinitetapes

21. Which is not the basic Regular expression pattern
    a. abc+
    b. a(a |b)
    c. abc?
    d. -abc-
22. Identify incorrect statement from following with respect Finite state automata (FSA) and Regular expression(RE)
    a. Regular expression is one way of describing FSA
    b. Both FSA and RE can be used to describe regular language
    c. Any FSA can be described with regular expression
    d. Any FSA regular expression cannot be implemented as FSA

23. Identify incorrect statement with respect to stochastic tagging

a. It requires vast amount of stored data

b. It is more popular than other tagging

c. In Rule based tagging is less accurate

d. Relatively complex

24. Which one is not aspect of pragmatics?

a. Deixis

b.Implicature

c. Reference Resolution

d. Presupposition

25. How many steps of NLP is there?

   a. 3
   b. 4
   c. 5
   d. 6

26. Which one is not the type of POS tagging

a. Rule based POS tagging

b. Stochastic POS tagging

c. Transformation based tagging

d. Image based tagging

27. A context free grammar is a _____

a. English grammar
b. Regular grammar
c. Context sensitive grammar
d. Left linear grammar

28. NLTK stands for ------------- and used for

a. Natural Language Toolkit

b. Natural Language Technique

c. Natural Language Technology

d. Natural Language Tokenization Kit

2 Marks Sample questions

29. Let G = (V, T, S, P) be a context-free grammar such that
Variables V = {S, R},
Terminal symbols T = {0, 1}
Productions P = {S → R1R1R1R, R → 0R | 1R |□ }
Which of the following languages are supported by this grammar?

a. L = {w | w contains at least three 1's}

b. L = {w | the length of w is odd and its middle is 0}

c. L = {w | w contains more 1's than 0's}

d. L = {w | w contains more 0's than 1's}

30. Which of the following instances the regular expression "\b(one|two|three)\b" can recognize?
a. one

b. onetwo

c. TWO

d. Three

31. Which of the following is an NLP task that involves determining all referring expressions that point to the same real-world entity?

a. Co reference resolution

b. Named entity recognition

c. Information extraction

d. Information Retrieval

32. ---------------- is a word with the most specific meaning.(U)

a. Homonym
b. Polysemy
c. Hyponymy
d. Synonymy

33. The word ------------ defines the relationship between different words that have a similar meaning.

a. Homonym
b. Polysemy
c. Hyponymy
d. Synonymy

34. Co reference Resolution is (U) *

   a. Anaphora Resolution
   b. Given a sentence or larger chunk of text, determine which words refer to the same objects
   c. Resolving Reference
   d. Ambiguity in Words

35. Text summarization is ------- learning approach
   a. Unsupervised
   b. Supervised
   c. Semi supervised
   d. Reinforcement

36. Following is not the variants of information retrieval system
   a. Bi-lingual Information Retrieval
   b. Cross-lingual Information Retrieval
   c. Half- lingual Information Retrieval
   d. Multi-lingual Information Retrieval

37. Name Entity Recognition is sub task of --------- that seeks to locate and classify named entities in texts.
   a. Information Retrieval
   b. Information Extraction
   c. Text summarization
   d. Text Classification

38. In linguistic morphology _____ is the process for reducing inflected words to their root form.

   a. Rooting
   b. b. Stemming
   c. Text-Proofing
   d. Lemmatization

39. "He lifted the beetle with red cap." contain which type of ambiguity?

a. Lexical ambiguity
b. Syntax Level ambiguity
c. Referential ambiguity
 d. Syntactical ambiguity

40.. Variation of word without changing its form is known as ----

a. Zero morphemes
b. One morphemes
c. Two morphemes
d. Three morphemes

41. Which of the following can be used to implement orthographic rules (Spelling Rules)?
a. Finite State Automata (FSA)
b. Finite State Transducer (FST)
c. Hidden Markov Model (HMM)
d. Turing Machine (TM)

42. What do we put after a character to match strings where that character appears two to four times in sequence?

a. {2,4}

b. {2-4}

c. [2,4]

d. <2-4>

43. The regular expression \d{4} will match what?

a. Any four character sequence?

b. Any four digit sequence?

c. The letter d four times?

d. Four time d?

44. In an HMM, observation likelihoods measure

a. The likelihood of a POS tag given a word

b. The likelihood of a POS tag given the preceding tag

c. The likelihood of a word given a POS tag

d. The likelihood of a POS tag given two preceding tags

45. "He was running quickly into the stadium". What type of phrase is this?

a. Noun phrase

b. Verb phrase

c. Prepositional phrase

d. Adjectival phrase

46. Which of the following is NOT the set of regular expression R = (ab + abb)* bbab

a.ababbbbab

b.abbbab

c.ababbabbbab

d.abababab

47. The study of the sound patterns in natural language and the rules that govern them is:

a. Phonetics

b. Morphology

c. Phonology

d. Syntax

48. Which of the following is not a part of speech?

a. Noun phrase

b. Verb

c. Interjection

d. Determiner

49. What is one advantage of a Porter stemmer over a full morphological analyzer?

 a. The stemmer is better justified from a theoretical point of view.

b. The stemmer does not require a detailed lexicon to implement.

 c. The output of the stemmer results in better performance in most downstream tasks, compared to the output of the morphological analyzer.

d. The output of the stemmer is always a valid word, unlike the output of the morphological analyzer.

50. A Hidden Markov Model is a (1), which means that the model is specified by (2). (Select the response that correctly completes both missing pieces.)

a. (1) discriminative model (2) P(X, Y)

b. (1) discriminative model (2) P(Y |X)

c. (1) generative model (2) P(X, Y)

d. (1) generative model (2) P(Y |X)

51. The words room and house are in a lexical semantic relation, in which room is the (1) and house is the (2).

a. (1) hypernym (2) hyponym

b. (1) hyponym (2) hypernym

c. (1) holonym (2) meronym

d. (1) meronym (2) holonym

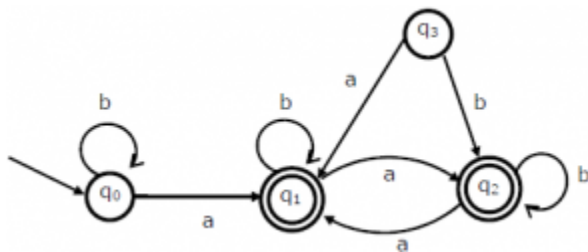52. Given a NFA with N states, the maximum number of states in an equivalent minimized DFA is at least

a. N^2

b. 2^N

c. 2N

d. N!

53. The language accepted by this DFA is

a. b*ab*ab*ab*

b. (a+b)*

c. b*a(a+b)*

d. b*ab*ab*

54. Which of the following is the instance of stemming as per the porter stemming algorithm?

a. are-> be

b. Plays->play

c. saw->s

d. university->universe

55. Which of following doesn't require application of NLP

a. Spam emails classification

b. Image classification of scanned handwritten documents

c. Generating captions for images

d. Sentiment analyizer for tweets

56. Identify incorrect statement about Finite state Transducer (FST) from the following

a. It is used as string recognizer

b. . It is used as string generator

c. It is used as translator

d. It is to covert string into machine language

57. Identify incorrect statement about Penn Treebank tag set

a. CD- Cardinal Data

b. VB- Verb Base form

c. VBD- Verb past Tense

d. NN-Noun

58. Which of the below are NLP use cases?
a. Detecting objects from an image
b. Facial Recognition
c. Speech Biometric
d. Question- Answering system

59. Identify the correct regular expression for given Automata
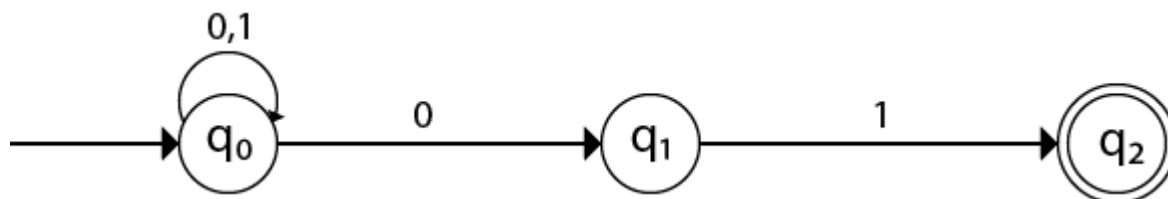
a. $\sum = \{0, 1\}$ accepts all string ending with 01.
b. $\sum = \{0, 1\}$ accepts all string in which the third symbol from the right end is always 0.
c. $\sum = \{0, 1\}$ in which double '1' is followed by double '0'.
d. $\sum = \{0, 1\}$ in which double '1' is followed by double '0' and '1'

60. What regular expression matches the whole words dog or cat?

    e. a. \bdog|cat\b

    f. b. \bdog\b | \bcat\b

    g. c. \bdog\b|\bcat\b

    h. d. \bdog\b|\bcat\b\bdog\bcat