

CENTRAL INSTITUTE OF TECHNOLOGY KOKRAJHAR
(Deemed to be University)
KOKRAJHAR :: BTR :: ASSAM :: 783370

END – SEMESTER EXAMINATION
UG

Session: Jan-Jun, 2023

Semester: VI

Time: 3Hrs.

Full Marks: 100

Course Code: UCSE613

Course Title: Speech & Natural Language Processing

Question 1 is compulsory answer any 3 from the rest!

1 [A] Objective Type Questions (MCQ) (1x 20)

1 x 20 = 20

- i. What is the syntax of Hindi Language?
 - A. SVO
 - B. VSO
 - C. SOV
 - D. OSV
- ii. Water can be written as water पानी, जल, नीर; Its an example of
 - A. Language Divergence
 - B. Language Characteristics
 - C. Ambiguity
 - D. None of the given
- iii. NLP is difficult for which of the following reason
 - A. Ambiguity
 - B. Word Order
 - C. Morphologically Richness
 - D. All of the given
- iv. Which one is not a Data Driven Machine Translation technique
 - A. Example based
 - B. Interlingua based
 - C. Neural based
 - D. Statistical based
- v. Which of the following is not necessary with respect to Ruled based Machine Translation
 - A. Required linguistic expertise to develop systems
 - B. Maintenance of system is difficult
 - C. Scaling to a large number of language pairs is not easy
 - D. Required Sophisticated High End Machine
- vi. How many word alignments are possible in between English: {Word1, Word2} and Hindi: {Word3, Word4}
 - A. 1
 - B. 2
 - C. 3

- D. 4
- vii. What is a monolingual corpus?
- A. Text of several languages
 - B. Text aligned for two languages
 - C. Text of a single language
 - D. None of the given
- viii. What is transliteration?
- A. Convert a text from one script to another script
 - B. Convert a text from one language to another language
 - C. Convert a text from other language
 - D. None of the given
- ix. What are the characteristics of related languages
- A. share significant vocabulary (cognates & loanwords)
 - B. correspondence between suffixes/post-positions
 - C. share the same basic word order
 - D. All of the given
- x. How do you measure the quality of a translation
- A. BLEU score
 - B. TER
 - C. METEOR
 - D. All of the given
- xi. Lexical semantics may be represented by
- A. Words Synonymy
 - B. Words Antonymy
 - C. Words Similarity
 - D. All of the above
- xii. What is the full form of TF-IDF?
- A. Term Frequency - Inverse Document Frequency
 - B. Time Frame – Inverse Document Frame
 - C. Term Finance – Intra Day Finance
 - D. Terminal Frequency – Inverse Document Frequency
- xiii. Which one is not a technique for generating Short Dense Vectors?
- A. Singular Value Decomposition
 - B. Brown Clustering
 - C. TF-IDF
 - D. Neural Language Model
- xiv. Skip-gram technique uses
- A. Target word to predict the context words
 - B. Context words to predict the target word
 - C. Context words to predict context words
 - D. None of the given
- xv. CBoW technique uses
- A. Target word to predict the context words
 - B. Context words to predict target word

- C. Target word to predict target word
 D. None of the given
- xvi. In NLP PoS stands for
 A. Point of Sale
 B. Parts of Speech
 C. Pre Order Sale
 D. None of the given
- xvii. Why PoS is useful?
 A. During Parsing need to know a word is Noun or Verb
 B. Finding the names in information extraction
 C. Speech synthesis/recognition
 D. All of the given
- xviii. What is the significance of a Language Modelling?
 A. Understand qualitative information
 B. Ability to learn from context
 C. Ability to predict what word comes next
 D. All of the given
- xix. Which one is not a Language Model
 A. Word2Vec
 B. EIMo
 C. BERT
 D. Hidden Markov Model
- xx. Emotion of a person can be detected using which of the following NLP Technique
 A. Word Sense Disambiguation
 B. Parts of Speech Tagging
 C. Sentiment Analysis
 D. None of the given

2 x 10 = 20

[B] Fill in the blanks (2x10)

- i. _____ the use of various statistical and probabilistic techniques to determine the probability of a given sequence of words occurring in a sentence.
- ii. The lower left corner and lower right corner there is _____ and _____ in Vauquois triangle.
- iii. A POS is a grammatical classification that commonly includes tags _____, _____, _____, _____ etc.
- iv. _____ is a natural language processing (NLP) technique used to determine whether data is positive, negative or neutral.
- v. CBoW in NLP stands for _____.
- vi. Skip-gram is a technique used in _____.
- vii. WSD in NLP stands for _____.
- viii. Statistical Machine Translation often consists of _____ and _____.

- ix. _____ and _____ are two frameworks for NLP
- x. Neural Language Modelling is a technique for NLP by the use of _____
- 2
- i. What do you understand by NLP? 2 + 8 + 10 = 20
 - ii. What are at least eight reasons that explain the challenges of Natural Language Processing (NLP)?
 - iii. Write at least ten applications of NLP.
- 3
- i. What is a Machine Translation? 2 + 8 + 10 = 20
 - ii. Explain the classification of Machine Translation.
 - iii. Explain Statistical Machine Translation in detail.
- 4
- i. What are vector semantics? 2 + 3 + 5 + 10
 - ii. What are the applications of vector semantics? = 20
 - iii. What are the different technique for vector semantics?
 - iv. Explain Word2Vec.
- 5
- i. What do you understand by PoS tagger and what is the importance of it? 6 + 6 + 8 = 20
 - ii. Briefly explain the applications of PoS Tagger.
 - iii. Explain the process of developing a PoS Tagger.
- 6 Write short notes (any four) 4 x 5 = 20
- a. Word Sense Disambiguation
 - b. Neural Machine Translation
 - c. Syntax Tree
 - d. Sentiment Analysis
 - e. Neural Language Modelling
 - f. BLEU