**WORKSHEET** **-3**

**NLP**

**All the question in this worksheet have one or more than one correct answers. Choose all the correct options to answer your question.**

1. Which of the following reduce a word to its base form by cutting off the suffix?

B) Porter Stemmer C) Snowball Stemmer D) WordNetLemmatizer

1. We need to perform stemming and lemmatization so that:
   1. All the words can be reduced to their base form
   2. so that we do not end up with too many words in the vocabulary which are not adding information to the model.
   3. so that lengths of words are reduced.

1. Stemming and Lemmatization belongs to which of the following step in NLP?

B) Syntactic Processing

C) Lexical Processing

1. Which of the following is/are example of shallow parsing?
   1. POS tagging B) Chunking
2. Which of the following are true regarding Lexicon Based taggers?
   1. These taggers assign that POS tag to the word whose frequency is maximum for that word in the training Corpus.
   2. These taggers also use tag of the previous word to find the tag of the word.
   3. These taggers take in to account the context in which the word is used to assign a tag.
   4. All of the above
3. Which of the following taggers uses predefined rules to assign tags?

B) Stochastic Rule taggers

C) Rule Based Taggers

1. Which of the following is /are true regarding HMM based POS tagger?
   1. It uses tag of only the previous word to determine the tag of the current word.
   2. It assigns tag by finding the most frequent tag occurring for that word in the training corpora
2. What does the transition probability refer to in HMM based POS tagging algorithm?
   1. The transition probabilities refer to probabilities of transitioning from one tag to another tag.
   2. Transition probabilities refer to the probability of emitting a given word from a tag.
   3. Transition probabilities are the probabilities of most occurring tag.

1. Which of the following are terminal symbols in the following Context-Free Grammar?

S -> NP VP

NP -> DT N| N| N PP VP -> V| V NP N -> ‘man’| ‘bear’

V -> ‘ate’

DT -> ‘the’| ‘a’

|  |  |
| --- | --- |
|  | B) ‘ate’ |
| C) VP |  |

1. In which of the cases Hidden Markov Model can be used?

B) POS tagging

C) Word Tokenization

1. Which of the following is/are used to get the grammatical construction of the sentence?

B) Constituency Parsing C) Top-Down Parsing

1. Which of the following are the approaches of constituency parsing?

C) Dependency Parsing

1. Which of the following is true regarding Top-Down parsing?

A) It starts with start symbol S.

* 1. we use the CFG production rule to generate the sentence from the S start symbol.

1. Which of the following statements are true regarding shift reduce parser algorithm?

A) It’s an algorithm of Bottom up parsing.

B) In this algorithm we start from the sentence, take one word at a time from the sentence shift it to the stack or reduce the words present in the stack by using CFG rules, until we reach the S start symbol. D) All of the above

1. Which of the following are true regarding Chomsky Normal Form?

A) It is normalized form of a CFG.

B) A CFG with no terminal symbol is called Chomksy Normal Form.

C)It is used for POS tagging.

1. In Which of the following text processing technique we will remove stopwords as a preprocessing?
   1. Top-Down Parsing B) Bottom-Up parsing
   2. Count-vectorization to create BOW for lexical level analysis.
   3. All of the above