NLP – WORKSHEET 4

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 are true regarding Context Free Grammars?
   A) It consists of a set of production rules
   B) The production rules are of the following form: A-> BC where A is non terminal while B, C can be either terminal or non-terminal
   C) These grammars are free of context in which they are used, they will remain same regardless of the context in which they are used.
   D) None of the above

   A) It consists of a set of production rules
   B) The production rules are of the following form: A-> BC where A is non terminal while B,

2. Advantages of using PCFG over CFG are:
   A) All the production rules in PCFG has probability associated with them while in CFG we do not have Probability of a production rule.
   B) With PCFG we can find the most probable parse tree of a sentence which we cannot find CFG.
   C) PCFG do not use probabilities while CFG uses probability in its production rules.
   D) All of the above

   A) All the production rules in PCFG has probability associated with them while in CFG we do not have Probability of a production rule.
   B) With PCFG we can find the most probable parse tree of a sentence which we cannot find CFG.

3. The problem with constituent parsing is:
   A) The constituent parsing is more complex
   B) The constituent parsing does not work with free word order languages where same meaning can be depicted with different word order.
   C) For free word order languages we cannot have a fixed set of production rules.
   D) All of the above

   D) All of the above

4. In order to deal with free word order languages what type of parsing is suitable?
   A) Free word parsing
   B) Constituency parsing
   C) Dependency Parsing
   D) None of the above

   A) Free word parsing

5. Which of the following are true regarding Dependency Parsing?
   A) It establish dependencies between words of a sentence
   B) The dependencies are established in terms of subject-object-verb and other dependencies.
   C) We make the parse tree in Top-Down approach
   D) None of the above

A) It establish dependencies between words of a sentence
B) The dependencies are established in terms of subject-object-verb and other dependencies.
C) We make the parse tree in Top-Down approach

6. Which of the following are techniques for Named Entity Recognition?
   A) Chunking B) Stochastic Rule taggers C) unigram chunker D) bigram chunker
   A) Chunking

7. Which of the following statements are true regarding Unigram chunker for NER?
   A) It uses the POS tag of a word and find the most probable IOB label for that POS tag
   B) It uses tag of only the previous word to determine the IOB label
   C) It uses the POS tags of the required word as well as previous word to assign IOB label
   D) None of the above

   C) It uses the POS tags of the required word as well as previous word to assign IOB label

8. Which of the following statements are true regarding Unigram chunker for NER?
   A) It assigns POS Tags to the words in a sentence
   B) It assigns IOB entity label to a word based on the POS tag of the word
   C) It assigns that IOB label which has the maximum probability based on the POS tag
   D) It uses the HMM model
   B) It assigns IOB entity label to a word based on the POS tag of the word

9. Which of the following statements are true regarding Bigram chunker for NER?
   A) It is a sequential modeling process for assigning POS tags to the word
   B) It uses a dictionary of IOB labels to assign a IOB entity label
   C) It uses the POS tag of the word and its previous word to assign the most probable IOB label
   D) The IOB label which occurs most frequently for a given pair of POS tags is assigned

   C) It uses the POS tag of the word and its previous word to assign the most probable IOB label

10. Which of the following technique uses a dictionary to extract an entity?
    A) Unigram chunker
    B) Bigram chunker
    C) Rule based chunking
    D) HMM based POS tagging

    A) Unigram chunker
    B) Bigram chunker
    C) Rule based chunking

11. Which of the following are preprocessing steps in Information Extraction Systems?
    A) Word Tokenization B) Lemmatization C) Dependency Parsing D) POS tagging

    A) Word Tokenization
    B) Lemmatization
    D) POS tagging

12. Which of the following cases require NER (Named Entity Recognition)?
A) POS Tagging
B) Chat Bot creation
C) A flight booking system which books flights for a customer according to information given by him in query
D) All of the above

D) All of the above

13. Which of the following is true regarding Top-Down parsing?
A) It starts with start symbol S
B) we use the CFG production rule to generate the sentence from the S start symbol
C) It starts with sentence and then we reduce it to the S symbol
D) All of the above

A) It starts with start symbol S
B) we use the CFG production rule to generate the sentence from the S start symbol

14. Consider the following string and tell what should be the pattern to extract all #tags String = "there should be justice for #sushant #singh @rindia" Import re Re.search(pattern, string)
A) pattern = "#\w*" B) pattern = "#*" C) pattern = "##" D) pattern = "#?"

A) pattern = "#\w*" B) pattern = "#*"

15. Consider the following string and tell what should be the pattern to extract the mention (@rindia) only String = "there should be justice for #sushant #singh @rindia" Import re Re.search(pattern, string)
A) pattern = "@*\w*" B) pattern = "@!:" C) pattern = "@\w*" D) pattern = "@?"

D) pattern = "@?"