**Natural Language Processing (17IT4704A)**

**A.Y 2022-23**

**1 Mark Questions**

**UNIT I**

**1. Define lexicon**

The list of stems and affixes, together with basic information about them (whether a stem is a Noun stem or a Verb stem, etc.)

**2. Define Morphological Parsing?**

It is the process of finding the **constituent morphemes** in a word (Eg. cat +N +PL for cats)

**3. What is disjunction operation?**

Disjunctionoperator, also called pipesymbol is **|**. The pattern **/cat|dog/** matches either the string cat or the string dog.

**4. List out regular languages supported operations.**

Intersection, Union, difference, complementation, reversal.

**5. Define FST?**

**Finite State Transducer** is a type of finite automaton which **maps between two sets of symbols**.

**6. What are Orthographic Rules?**

These are spelling rules used to model the changes that occur in a word, when two morphemes combine.

**7. Write the regular expression to specify both guppy and guppies?**

/guppy|ies/

**8. Define Derivation?**

It is the combination of a word stem with a grammatical morpheme,usually resulting in a word of different

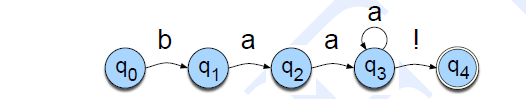
**9. What is operator precedency hierarchy?**

Highest to Lowest Precedence: Paranthesis (),Counters \* + ?{} ,Sequences and anchors the ^my end, Disjunction | .

**10. Write regular expression for the set of all lowercase alphabetic strings ending with b.**

[a-z] \* b

**11. Draw a finite-state automaton for talking sheep.**



What are the three equivalent ways of describing regular languages?

Define Inflection?

12.**Write short notes on lexicography.**

The identification of a string of words as a lexical unit, the basic unit of meaning.

**UNIT II**

**1. What is tagging?**

Process of assigning a part-of-speech to each word in a corpus

**2. Write the chain rule of probability to words.**

P(X1...Xn) = P(X1)P(X2|X1)P(X3|X12)...P(Xn|X1 n−1) = n ∏ k=1 P(Xk|X1 k−1 )

**3. What are auxiliaries?**

The subclass of verbs called **auxiliaries** or **helping verbs.** Auxiliaries include the modal verbs *can, could, may, might, must, will, would, shall*, and *should.*

**4. Define Nominalization.**

The process of forming a noun or noun phrase from a clause or a verb.

**5. What is Laplace smoothing?**

Take our matrix of bigram counts, before we normalize them into probabilities; **add one to all the counts**. This algorithm is called Laplace smoothing.

**6. What is stochastic part of speech tagging?**

Stochastic taggers use **probabilistic and statistical information** to assign **tags to words**. These taggers might use ‘tag sequence probabilities’, ‘word frequency measurements’ or a combination of both.

**7. How can we estimate N-gram probabilities in simplest way?**

The simplest and most intuitive way to estimate probabilities is called **Maximum Likelihood Estimation**, **MLE**. We get the MLE estimate for the parameters of an *N*-gram model by taking counts from a corpus, and **normalizing** them so they lie between 0 and 1.3

**8. Define closed class types and open class types.**

Closed classes are those that have relatively fixed membership. Eg Prepositions. Open classes are those that don’t have fixed membership. Nouns and verbs are open classes because new nouns and verbs are continually coined from other languages.

**9. What is interpolation?**

Mixing the probability estimates from all the *N*-gram estimators, i.e., we do a weighted interpolation of trigram, bigram, and unigram counts.

**10. What is tagset of English?**

**11.Define Markov Chains?**

**12.Define Markov Model?**

**UNIT III**

**1. What is parsing and list out parsing techniques?**

Parsing is a combination of recognizing an input string and assigning a structure to it. Parsing Techniques are Top down and Bottom up parsing.

**2. Define context-free grammar**

The most commonly used mathematical system for modeling constituent structure in CFG English and other natural languages is the Context-Free Grammar or CFG.

**3. Define ambiguity in context free grammar**

Assigning more than one possible parse tree to a sentence by the grammar.

**4. What is thematic role?**

The study of roles associated with specific verbs and across classes of verbs is usually referred to as thematic role or case role.

**5. Define generative grammar?**

The use of formal languages to model NLs is called **generative grammar**, since the language is defined by the set of possible sentences “generated” by the grammar.

**6. Define Sentences with declarative structure**

A subject NP followed by a VP

**7. Define Sentence with imperative structure**

Begin with a VP and have no subject.

**8. Define Sentences with yes-no-question structure**

Begin with auxiliary, followed by a subject *NP*, followed by a *VP*

**9. Define Sentences wh-subject-question structure**

Identical to the declarative structure, except that the first NP contains some **wh-word**

**10. What are Pre determiners?**

Word classes appearing in the NP before the determiner Eg:-***all*** *the flights,* ***all*** *flights*

**11. What are Post determiners?**

Word classes appearing in the NP between the determiner and the head noun

**12. Define** **adjective phrase?**

Adjectives can be grouped into a phrase called an adjective phrase or AP.

AP can have an adverb before the adjective - *the* ***least*** *expensive fare*

**13. What is gerundive verb?**

A gerundive consists of a VP begins with the gerundive (*-ing*)

**14. What are** **coordinations?**

*and, or,* and *but.*

**15. Define Equivalent of Grammars?**

Two grammars are equivalent if they generate the same set of strings.

**16. Define** **constituent?**

Groups of words may behave as a single unit or phrase, called constituent, e.g., NP

**17. Differentiate between top down and bottom up parsing?**

The top-down strategy never wastes time exploring trees that cannot result in an *S*.

The bottom-up strategy trees that have no hope to leading to an *S*, or fitting in with any of their neighbors

**18. Define strong equivalent grammars?**

If two grammars generate the same set of strings *and* they assign the same phrase structure to each sentence

**19. Define weakly equivalent grammars?**

Two grammars generate the same set of strings but do not assign the same phrase structure to each sentence.

**20. Define CNF?**

**Chomsky normal form** (CNF) it is ε-free and in addition each production is either of the form ***A* → *B C*** or ***A* → *a***

**21. What are the methods for dealing with left-recursion?**

Rewriting the grammar

*A* → *A*β | α ⇒ *A* → α *A’*

*A’* → β *A* | ε

**22. Define Binary branching**

It is a constraint on concatenative word formation which says that in the process of word formation only two morphemes can be concatenated at the same time.

**23. What are different types of ambiguities?**

Structural ambiguity (Attachment, Coordination), Syntactic ambiguity, Semantic ambiguity.

**UNIT IV**

**1. Define syntax and give suitable example**

The word syntaxmeaning “setting out together or arrangement”, and refers to the way words are arranged together. Eg: *On September seventeenth*, I’d like to fly from Atlanta to Denver

**2. What is semantic analyzer?**

Ensuring that the declarations and statements of a program are semantically correct, i.e, that their meaning is clear and consistent with the way in which control structures and data types are supposed to be used.

**3. Write the FOL representation for Maharani serves vegetarian food.**

Serves (Maharani, Vegetarian Food)

**4. Describe principle of compositionality**

The Principle of compositionality is the principle that the meaning of a complex expression is determined by the meanings of its constituent expressions and the rules used to combine them.

**5. Define Binary branching**

It is a constraint on concatenative word formation which says that in the process of word formation only two morphemes can be concatenated at the same time.

**6. Draw the diagram of pipeline approach to Semantic analysis**

