Unit I

S.NO	Question
1	Explain how the Stop word removal and Rare word removal are performed using NLTK with examples.
2	Compute the minimum edit distance between intention and execution using the minimum edit distance algorithm.
3	How Text wrangling and Cleansing are performed by using NLTK.
4	Explain in detail about basic regular expression patterns
5	Write regular expressions for the following languages. 1. the set of all alphabetic strings; 2. the set of all lower case alphabetic strings ending in a b; 3. the set of all strings from the alphabet a,b such that each a is immediately preceded by and immediately followed by a b 4. The set of all strings with two consecutive repeated words.
6	Explain Byte-Pair Encoding for Tokenization with an Example
7	What are the Unix Tools used for Crude Tokenization and Normalization?
8	Write a RE to find cases of the English article “the”.
9	Explain how Tokenization, Stemming, Lemmatization will be done using NLTK.

Unit II

S.NO	Question
1	Explain with an example on training and testing the naive Bayes with add-one smoothing
2	Illustrate Laplace smoothing with an example
3	Discuss Naive Bayes Classifiers?
4	We are given the following corpus <S> I am Sam</S> <S> Sam I am</S> <S> I am Sam</S> <S> I do not like green eggs and Sam</S> Using a bigram language model with add-one smoothing, what is P(Sam | am)? Include <S> and </S>in your counts just like any other token.
5	Compute the probability of sentences like I want English food or I want Chinese food in the Berkeley Restaurant Project corpus
6	Perplexity is the metric for evaluating language models. Discuss.
7	Write a short notes on and Interpolation.
8	Assume the following likelihoods for each word being part of a positive or negative movie review, and equal prior probabilities for each class. pos neg I 0.09 0.16 always 0.07 0.06 like 0.29 0.06 foreign 0.04 0.15 films 0.08 0.11 What class will Naive bayes assign to the sentence "I always like foreign films."?
9	Given the following short movie reviews, each labeled with a genre, either comedy or action: 1. fun, couple, love, love comedy 2. fast, furious, shoot action 3. couple, fly, fast, fun, fun comedy 4. furious, shoot, shoot, fun action 5. fly, fast, shoot, love action and a new document D: fast, couple, shoot, fly compute the most likely class for D.

	Assume a naive Bayes classifier and use add-1 smoothing for the likelihoods.
10	Discuss in detail about how to Train the Naive Bayes Classifier
11	Explain with an example about binarization for the binary naive Bayes algorithm
12	Discuss the difference between Laplace smoothing and Add-k smoothing.

Unit III

S.NO	Question
1	Parse the sentence “Book the flight through Houston” using CKY algorithm
2	Construct suitable example to discuss about ambiguity.
3	Identify the categories of English word classes and explain in detail
4	Describe Hidden markov model with suitable example
5	Find and Explain one tagging error in each of the following sentences that are tagged with the Penn Treebank tagset: 1. I/PRP need/VBP a/DT flight/NN from/IN Atlanta/NN 2. Does/VBZ this/DT flight/NN serve/VB dinner/NNS 3. I/PRP have/VB a/DT friend/NN living/VBG in/IN Denver/NNP 4. Can/VBP you/PRP list/VB the/DT nonstop/JJ afternoon/NN flights/NNS
6	Use the Penn Treebank tagset to tag each word in the following sentences from Damon Runyon’s short stories. You may ignore punctuation. 1.It is a nice night. 2. This crap game is over a garage in Fifty-second Street. . . 3. . . . Nobody ever takes the newspapers she sells . . .
7	Explain NER as a sequence model, which shows IO, BIO, and BIOES taggings.
8	List the Examples of type ambiguities in the use of the name Washington.
9	Write short notes on Named Entities and Named Entity Tagging.
10	Illustrate POS Tagging with suitable Examples.
11	Discuss about Penn Treebank part-of-speech tags.
12	Draw and Explain The parse tree for “I prefer a morning flight” according to grammar L0.
13	Discuss in detail about “Constituency”.
14	What is CFG?Discuss in detail
15	What is meant by Grammar Equivalence? How convert a given grammar into CNF?

Unit IV

S.NO	Question
1	Write a notes on the semantics of First-Order Logic
2	Discuss about modus ponens and explain how it is used in forward and backward chaining?
3	Illustrate Description Logics with suitable examples
4	Give FOL translations for the following sentences: 1. Vegetarians do not eat meat. 2. Not all vegetarians eat eggs.
5	Give a set of facts and inferences necessary to prove the following assertions: 1. McDonald's is not a vegetarian restaurant. 2. Some vegetarians can eat at McDonald's. Don't just place these facts in your knowledge base. Show that they can be inferred from some more general facts about vegetarians and McDonald's.
6	∀xVegetarianRestaurant(x) =⇒ Serves(x,VegetarianFood) Give a FOL rule that better defines vegetarian restaurants in terms of what they serve.
7	List Some of the noun and verb relations in WordNet
8	Discuss and explain about the Relations Between Senses
9	Define and discuss about Word Senses
10	Write the Hyponymy chains for two separate senses of the lemma **bass**
11	Write a short note on Unambiguous Representations and Canonical Form.
12	How to bridge the gap from formal representations to representations that tell us something about some state of affairs in the world? Explain?
13	What is Model-Theoretic Semantics? Discuss.
14	Write the FOL Representation for the following sentences.Explain I only have five dollars and I don't have a lot of time.

	if AyCaramba is a Mexican restaurant near ICSI All vegetarian restaurants serve vegetarian food.
15	Write the context-free grammar specification of the syntax of First-Order Logic representations.
16	Discuss in detail about lambda notation.
17	What is inference? Explain modus pones in detail?
18	Write a short notes on Event and State representations.
19	Write the detailed model for the restaurant world.