**NLP Question Bank**

**Unit I Introduction**

* Origins and challenges of NLP, limitations,
* Language Modeling - - Grammar-based LM, - Statistical LM- Differences
* Regular Expressions - various characters and their meaning and examples.
* Finite-State Automata – various parsing techniques
* Tokenization – process of tokenization, example, types
* Detecting and Correcting Spelling Errors – various techniques, Minimum Edit Distance, algo, example.

**Unit II Word Level Analysis**

* N Gram language modeling, their formulae.
* Part-of-Speech Tagging - examples
* Issues in PoS tagging
* Hidden Markov models – Introduction, Mathematics- formulae, examples, calculate best path given observations and emission data. Solve the following for any path. Example 3 1 3.
* Maximum Entropy models

**Unit III Syntactic Analysis**

* Context-Free Grammars
* Given a Grammer check whether it is valid or no using top to bottom and bottom up parsing.
* Treebanks
* Normal Forms for grammar – CFG to CNF conversion examples.
* Given a grammar, check whether the sentence is valid or not using CKY.
* Dependency Grammar
* Ambiguity
* Shallow parsing
* Viterbi algorithm and its role in finding the most likely tag sequence.

**Unit IV Semantics And Pragmatics**

* First-Order Logic - given statements write the FOL
* requirements for knowledge representation in NLP.
* First-Order Logic to represent semantic meaning.
* Thematic Roles (e.g., Patient, Instrument)- examples, applications.

**Unit V**

* Discourse Segmentation – challenges involved in segmenting a text into coherent discourse units.
* Word senses and importance
* Porter Stemmer, Lemmatizer, WordNet, Brown Corpus, Frame Net
* Coherence and coherence reference phenomena