NAME:SamprityGhosh

Roll: 21CS011119

Subject Name: NLP

Semester: 8th Sem

Topic: Assignment

1. a) Evolution of NLP from rule-based approaches to deep learning models.

NLP has evolved from rule-based systems using handcrafted linguistic rules to statist
methods relying on probabilistic models, and now to deep learning models that leverag
networks. Deep learning enables context-aware language processing, handling complex p
and large datasets, making NLP more accurate and versatile in applications like machin
translation, sentiment analysis, and chatbots.

Key Points:

Rule-Based Systems: Early NLP systems relied on predefined grammatical and syntactic
These systems were limited by their inability to handle ambiguity or variations in lang
Statistical Methods: Introduced probabilistic models like Hidden Markov Models (HMMs)
grams, which improved tasks like machine translation and speech recognition by learnin
from data.

Deep Learning: Modern NLP uses neural networks (e.g., RNNs, LSTMs, Transformers) to c
context and semantics. Models like BERT and GPT leverage large datasets for tasks like
analysis, text generation, and question answering.

Impact: Deep learning has revolutionized NLP, enabling more accurate, scalable, and con
language processing.

1. b) Different tokenization techniques used in lexical analysis.

Tokenization is the process of breaking text into smaller units like words, sentences, o
subwords. Techniques include whitespace tokenization, punctuation-based tokenization
subword tokenization. These methods are essential for preprocessing text in NLP task
efficient parsing, machine translation, and information retrieval by converting raw text
manageable and meaningful units.

Key Points :

Whitespace Tokenization: Splits text based on spaces, commonly used for languages lik
It is simple but struggles with punctuation and complex word structures.

Punctuation-Based Tokenization: Uses punctuation marks as delimiters, improving hand contractions and abbreviations.

Subword Tokenization: Breaks words into smaller units (e.g., Byte Pair Encoding, WordPi

for handling rare words and morphologically rich languages.

Sentence Tokenization: Splits text into sentences, often using punctuation and capitaliza

Applications: Tokenization is crucial for tasks like machine translation, sentiment analy

information retrieval, as it converts raw text into structured input for NLP models.

2. a) How does Part-of-Speech (POS) tagging contribute to NLP applications?

POS tagging assigns grammatical categories (e.g., noun, verb) to words in a sentence. It

NLP applications by improving syntax analysis, disambiguating word meanings, and aidin

like machine translation, information extraction, and text-to-speech systems. POS tagg

provides structural context, enabling more accurate and meaningful language processin

Key Points :

Syntax Analysis: POS tagging helps parse sentence structure, identifying subjects, object

predicates.

Disambiguation: Resolves word meanings based on context (e.g., "bank" as a noun or ve

Machine Translation: Improves translation accuracy by understanding grammatical role

Information Extraction: Aids in identifying entities and relationships in text.

Text-to-Speech: Enhances natural-sounding speech synthesis by providing grammatical

Techniques: Rule-based, statistical (e.g., HMMs), and deep learning-based (e.g., BiLSTM) me

used for POS tagging.

Applications: Used in chatbots, search engines, and sentiment analysis for better langua

understanding.

2. b) Process of semantic analysis, including word embedding and sentiment analysis.

Semantic analysis interprets the meaning of text by analyzing context and relationships

words. Techniques like word embedding (e.g., Word2Vec, GloVe) represent words in vecto

capturing semantic relationships. Sentiment analysis determines the emotional tone of

classifying it as positive, negative, or neutral. These techniques are vital for applicatio chatbots, recommendation systems, and opinion mining.

Key Points :

Word Embedding: Represents words as vectors in a continuous space, capturing semar
relationships (e.g., Word2Vec, GloVe).

Contextual Understanding: Models like BERT and GPT use transformers to capture conte
meaning.

Sentiment Analysis: Classifies text based on emotional tone, useful for brand monitori
customer feedback.

Applications: Used in chatbots, recommendation systems, and opinion mining.

Techniques: Includes lexicon-based methods, machine learning (e.g., SVM), and deep learni
RNNs, Transformers).

Challenges: Handling sarcasm, ambiguity, and cultural nuances in text.

Impact: Enhances user experience by enabling more accurate and context-aware langu
processing.