

NLP Assignment 5

1. Explain the concept of lemmatization.

Lemmatization is a linguistic process that involves reducing words to their base or root form, known as the lemma. It considers the context and part of speech of the word to provide a meaningful lemma. For example, lemmatizing the word "running" would result in "run" as the lemma.

2. What exactly is NLU? Mention how it's used.

NLU is a branch of artificial intelligence that focuses on enabling machines to comprehend and interpret human language in a valuable and meaningful manner. It involves tasks like sentiment analysis, intent recognition, and named entity recognition. NLU is used in chatbots, virtual assistants, and various applications to understand and respond to human queries.

3. What exactly are stop words?

Stop words are common words, such as "and," "the," "is," etc., that are often removed during text preprocessing in NLP. These words typically do not carry significant meaning and are frequently used in a language. Removing stop words can reduce noise and improve the efficiency of NLP tasks.

4. What is corpus juris?

"Corpus Juris" usually refers to a specialized collection of legal texts, laws, statutes, cases, regulations, and legal literature used for training and developing NLP models for legal domain-specific tasks. NLP researchers and developers often use such specialized corpora to build models that can understand legal language, extract legal entities, identify legal concepts, or even predict legal outcomes based on historical cases. The term "corpus juris" in this context essentially denotes a curated dataset of legal texts used for NLP applications within the legal domain.]

5. Describe the veiled language model.

The term "veiled language model" is not a standard concept in NLP or linguistics. If you meant a different term or have a specific context, feel free to provide more details for a more accurate explanation.

6. Explain how a word embedding works.

Word embedding is a technique that converts words or phrases into vectors (numerical representations) in a continuous vector space. Each word is mapped to a unique vector, and these vectors capture semantic relationships between words. Proximity or similarity in the vector space indicates similarity in meaning, enabling machine learning models to process and understand words in a more efficient and meaningful way.