Natural Language Processing
Lab 4

This lab sheet is to practice the concepts taught this week so far: Lexical Semantics.

1. What is synonymy, and why is it important in NLP?

   Ans: Synonymy refers to the relationship between words with similar meanings. In NLP, it's crucial for understanding user intent, enhancing language models, and improving tasks like search queries and content recommendation.

2. How do NLP systems identify synonyms?

   Ans: NLP systems identify synonyms using lexical databases like WordNet, machine learning models that analyze word usage patterns in large text corpora, and contextual embeddings from models like BERT (which will be discussed in future weeks).

3. Think about and discuss the challenges of synonymy for machine translation systems.

   Ans: In machine translation, synonymy presents challenges in choosing the right word that fits the context in the target language, affecting translation accuracy and fluency.

4. Think about and discuss the limitations of current NLP technologies in dealing with synonymy?

   Ans: Current limitations include difficulty in capturing all possible synonyms, especially in domain-specific contexts, and challenges in handling polysemy (words with multiple meanings).

5. Imagine you have word vectors for "king" = [0.1, 0.3, 0.7] and "queen" = [0.2, 0.4, 0.6]. Calculate the cosine similarity to determine how similar these words are in the vector space.

Ans:

dot product = 0.56
$\|king\| = 0.77$
$\|queen\| = 0.75$
$cosinesimilarity = 0.97$
$Verysimilar.$

6. In a corpus analysis focusing on word associations, you're investigating the association between the words "sunny" and "weather". Based on the corpus data, the probabilities are as follows:

   - $P(\text{"}sunny\text{"}) = 0.05$
   - $P(\text{"}weather\text{"}) = 0.1$
   - $P(\text{"}sunny\text{"}, \text{"}weather\text{"}) = 0.02$

   Calculate the Pointwise Mutual Information (PMI) between "sunny" and "weather". What does the PMI value indicate about the relationship between these two words in the corpus?

Ans: 1.39. A positive PMI result would indicate a strong association between "sunny" and "weather" in the corpus, suggesting that these words co-occur more frequently than by chance alone.