Natural Language Processing
Lab 4

February 5, 2024

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?
2. How do NLP systems identify synonyms?
3. Think about and discuss the challenges of synonymy for machine translation systems.
4. Think about and discuss the limitations of current NLP technologies in dealing with synonymy?
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.
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?