

## Week 8 Quiz

### Q1.

What makes natural languages complex for computer processing?

- A. Their binary structure
- B. The ambiguity and contextual nuances
- C. The lack of vocabulary
- D. Their universal grammar rules

### Q2.

What is a significant challenge in rule-based language models?

- A. They require too much electricity to run
- B. They can only understand English
- C. It is difficult to write rules for all possible sentences
- D. They are too easy for computers to process

### Q3.

The transition from rule-based to statistical models in NLP was primarily due to:

- A. The invention of the internet
- B. The availability of large-scale datasets and advancements in deep learning
- C. The decrease in interest in linguistics
- D. The need for faster email processing

### Q4.

In the context of NLP, what does tokenization achieve?

- A. Increases the size of text documents
- B. Breaks text into words, phrases, symbols, or other meaningful elements
- C. Translates text from one language to another
- D. Summarizes long documents

**Q5.**

What is an example of a task that uses Named Entity Recognition (NER)?

- A. Correcting spelling mistakes
- B. Counting the number of paragraphs in a text
- C. Identifying and classifying entities like persons and locations in text
- D. Changing the font size of a document

**Q6.**

Why are transformers considered more efficient than RNNs for processing long sequences of data?

- A. Because they can transform text into images
- B. Because they require less electricity to operate
- C. Because they can better handle long-range dependencies using the attention mechanism
- D. Because they are newer and thus inherently better

**Q7.**

How does lemmatization differ from stemming in natural language processing?

- A. Lemmatization is faster than stemming.
- B. Lemmatization considers the context and converts the word to its meaningful base form, whereas stemming simply removes suffixes.
- C. Stemming is used only for English, but lemmatization works for all languages.
- D. There's no significant difference; both processes produce identical results.

**Q8.**

What principle does the Word2Vec model primarily rely on?

- A. Words that are used in the same document share similar meanings.
- B. Words that appear in similar contexts tend to have similar meanings.
- C. Every word in a language can be represented by a unique number.
- D. The meaning of a word can be determined by its spelling.

**Q9.**

In the context of NLP, what is the purpose of creating a Document-Term Matrix (DTM)?

- A. To translate documents from one language to another.
- B. To summarize long documents into shorter ones.
- C. To represent text data numerically by counting the frequency of terms in documents, facilitating further analysis.
- D. To encrypt text data for secure communication.

### **Q10.**

Which of the following is a challenge when using Bag-of-Words (BoW) representation?

- A. It can accurately predict the next word in a sentence.
- B. It loses information about the order of words, which can be crucial for understanding the meaning.
- C. It is too complex for computers to process.
- D. It cannot be used for English language text.

### **Q11.**

What advantage do "large language models" have over smaller models?

- A. They are easier to store and distribute.
- B. They can capture a broader understanding of language nuances, syntax, and semantics due to being trained on extensive data.
- C. They always provide perfect translations between languages.
- D. They require less energy to run.