# Week 1 Practice Quiz

**10/10** points earned (100%)

Excellent!

Retake

[Course Home](https://www.coursera.org/learn/text-mining/home/welcome)

Correct

1 / 1 points

1. Which of the following word pairs are paradigmatically related? Check all that apply.

* **Car, vehicle**

**Correct Response**

Yes! This is one of the correct answers because the word pairs tend to occur in the same context. We would generally not obtain a meaningful sentence by substituting “car” for “drive.”

* **Computer, laptop**

**Correct Response**

Yes! This is one of the correct answers because the word pairs tend to occur in the same context. We would generally not obtain a meaningful sentence by substituting “laptop” for “keyboard.”

* **Car, drive**

**Correct Response**

* **Computer, keyboard**

**Correct Response**

Correct

1 / 1 points

2. Which of the following word pairs are syntagmatically related? Check all that apply.

* **Car, drive**

**Correct Response**

Yes! This is one of the correct answers because the words in this pair tend to co-occur in the same sentences.

* **Car, vehicle**

**Correct Response**

* **Computer, laptop**

**Correct Response**

* **Computer, keyboard**

**Correct Response**

Yes! This is one of the correct answers because the words in this pair tend to co-occur in the same sentences.

Correct

1 / 1 points

3. Suppose the pseudo-document representations for the contexts of the terms A and B in the vector space model are given as follows:

* dA = (0.30, 0.20, 0.40, 0.05, 0.00, 0.05)
* dB = (0.40, 0.10, 0.30, 0.00, 0.20, 0.00)

What is the EOWC similarity score?

1. 0.40
2. 0.22
3. 0.32
4. **0.26**

**Correct Response**

0.3 ∗ 0.4 + 0.2 ∗ 0.1 + 0.4 ∗ 0.3 = 0.26

Correct

1 / 1 points

4. True or false? EOWC treats words irrespective of their occurrence in the corpus. Adding IDF (invert document frequency) weighting to the EOWC similarity function will penalize common words in the corpus.

1. **True**

**Correct Response**

IDF favors rare words and penalize common words

1. False

Correct

1 / 1 points

5. True or false? The EOWC score is non-negative and cannot exceed 1.

1. False
2. **True**

**Correct Response**

The EOWC score is the probability that two randomly selected words from the two pseudo-documents are the same. Since it is a probability, its values range between 0 and 1 (inclusive).

Correct

1 / 1 points

6. Which of the following is not the reason that NLP is difficult?

1. Word-level ambiguity, such as the word "design" can be a noun or a verb
2. Coreference resolution, such as reference of "he," "she," or "it"
3. Syntactic ambiguity, such as PP attachment
4. **Difficulty of transforming non-text data into text data**

**Correct Response**

This is not the main issues of NLP

Correct

1 / 1 points

7. Which of the following statements about NLP is correct?

1. Summarization and translation are considered to be shallow NLP.
2. **Shallow NLP tends to be robust, while deep NLP doesn't scale up well.**

**Correct Response**

deep NLP are much more complicated and difficult

1. POS-tagging is considered to be deep NLP.
2. For English, parsing POS-tagging with 100% accuracy is possible if computing resource and time is unlimited.

Correct

1 / 1 points

8. For sentiment analysis (opinion mining), which kind of text representation is generally used?

1. **Words, syntactic structures, entities, and relations**

**Correct Response**

1. Words
2. String
3. Words and syntactic structures

Correct

1 / 1 points

9. Paradigmatic related words have:

1. **High context similarity**

**Correct Response**

such as Monday, Tuesday, Wednesday ...

1. High co-occurrence

Correct

1 / 1 points

10. Which of the following help to improve EOWC? Check all that apply.

* **Reward matching rare words and discount matching frequent ones**

**Correct Response**

This favors rare words

* **Use sublinear transformation of term frequency**

**Correct Response**

The increase of EOWC is smaller for higher TF