Distributed Information Systems: Spring Semester 2018
Quiz 6 - May 31 2018

1. Which of the following is wrong regarding *Ontologies*?
   - ☐ a. We can create more than one ontology that conceptualize the same real-world entities
   - ☐ b. Ontologies help in the integration of data expressed in different models
   - ☐ c. Ontologies support domain-specific vocabularies
   - ☐ d. Ontologies dictate how semi-structured data are serialized

2. Which of the following is correct regarding the use of *Hidden Markov Models (HMMs)* for entity recognition in text documents?
   - ☐ a. HMMs cannot predict the label of a word that appears only in the test set
   - ☐ b. If the smoothing parameter λ is equal to *1*, the emission probabilities for all the words in the test set will be equal
   - ☐ c. When computing the emission probabilities, a word can be replaced by a morphological feature (e.g., the number of uppercase first characters)
   - ☐ d. The label of one word is predicted based on all the previous labels

3. Which of the following statements is correct concerning the use of *Pearson's Correlation* for user-based collaborative filtering?
   - ☐ a. It measures whether different users have similar preferences for the same items
   - ☐ b. It measures how much a user's ratings deviate from the average ratings
   - ☐ c. It measures how well the recommendations match the user's preferences
   - ☐ d. It measures whether a user has similar preferences for different items

4. Which of the following statements is wrong regarding RDF?
   - ☐ a. An RDF statement would be expressed in SQL as a tuple in a table
   - ☐ b. Blank nodes in RDF graphs correspond to the special value NULL in SQL
   - ☐ c. The object value of a type statement corresponds to a table name in SQL
   - ☐ d. RDF graphs can be encoded as SQL databases

5. Which of the following statements is correct in the context of information extraction?
   - ☐ a. A confidence measure that prunes too permissive patterns discovered with bootstrapping can help reducing semantic drift
   - ☐ b. The bootstrapping technique requires a dataset where statements are labelled
   - ☐ c. Distant supervision typically uses low-complexity features only, due to the lack of training data
   - ☐ d. For supervised learning, sentences in which NER has detected no entities are used as negative samples

6. Which of the following is correct regarding schemas and ontologies?
   - ☐ a. An ontology is created from constructing mappings between schemas
   - ☐ b. Ontologies can be used for reasoning about different schemas
   - ☐ c. Ontologies always require a schema

☐ d.   Semi-structured data cannot have a schema

7. *Dude said "I like bowling"*. With how many statements can we express this sentence using *RDF Reification*?
    ☐ a.   We cannot
    ☐ b.   1
    ☐ c.3
    ☐ d.   5

8. Modularity of a social network always:
    ☐ a.   Increases with the number of communities
    ☐ b.   Increases when an edge is added between two members of the same community
    ☐ c.Decreases when new nodes are added to the social network that form their own communities
    ☐ d.   Decreases if an edge is removed