Distributed Information Systems: Spring Semester 2018 - Quiz 1

Student Name: _____
Date: May 18 2018
Student ID: _____
Total number of questions: 8
Each question has a single answer!

_____

1. Data being classified as unstructured or structured depends on the:
    A. Degree of abstraction
    B. Level of human involvement
    C. Type of physical storage
    D. Amount of data

2. Which of the following is an advantage of Vector Space Retrieval model?
    A. No theoretical justification is needed why the model works
    B. Produces provably correct query results
    C. Enables ranking of query results according to cosine similarity function
    D. Allows to retrieve documents that do not contain any of the query terms

3. Which of the following is *true*?
    A. High precision implies low recall
    B. High precision hurts recall
    C. High recall hurts precision
    D. High recall implies low precisions

4. Recall can be defined as:
    A. P(relevant documents | retrieved documents)
    B. P(retrieved documents | relevant documents)
    C. P(retrieved documents | number of documents)
    D. P(relevant documents | number of documents)

5. Thang, Jeremie and Tugrulcan have built their own search engines. For a query Q, they got precision scores of 0.6, 0.7, 0.8 respectively. Their F1 scores (calculated by same parameters) are same. Whose search engine has a higher recall on Q?
    A. Thang
    B. Jeremie
    C. Tugrulcan
    D. We need more information

6. The number of non-zero entries in a column of a term-document matrix indicates:
    A. how many terms of the vocabulary a document contains
    B. how often a term of the vocabulary occurs in a document
    C. how relevant a term is for a document
    D. none of the other responses is correct

7. Which one of the following is *wrong*. Schema mapping is used to:
    A. Overcome semantic heterogeneity
    B. Reconcile different logical representations of the same domain
    C. Optimize the processing of queries
    D. Support schema evolution of databases

8. In a Ranked Retrieval result, the result at position k is non-relevant and at k+1 is relevant. Which of the following is *always* true (P@k and R@k are the precision and recall of the result set consisting of the k top ranked documents)?
    A. P@k-1 > P@k+1
    B. P@k-1 = P@k+1
    C. R@k-1 < R@k+
    D. R@k-1 = R@k+1