

**Questions and Answers**

1. By creating an index the data will increase in term of size:  
☒ True.  
☐ False.
2. Unstructured data means:  
☐ Has clear semantic forms.  
☒ Has not clear semantic forms.  
☐ None of the above.
3. Information retrieval goal is:  
☐ Build indices for data.  
☐ Retrieve largest document in the collection.  
☐ Retrieve non-relevant documents to a user query.  
☒ None of the above.
4. Information retrieval match query directly to document data (ex: PDF file):  
☐ True.  
☒ False.
5. Does best match IR model type, result a ranked list of documents (best result is the first):  
☒ True  
☐ False
6. Exact matching may retrieve too few or too many documents:  
☒ True  
☐ False
7. In what situation rare terms have the same value of frequent terms:  
☐ All  $tf_w$  of terms = 1.  
☐ All  $tf_w$  of terms = 0.  
☐ All terms have the same  $idf = tf_w$ .  
☐ All  $idf$  of terms = 0.  
☒ All  $idf$  of terms = 1.
8. CRISP-DM methodology that provide blueprint for IR system:  
☐ True.  
☒ False.
9. Assume that the confidence of the rule 1->2 is 100%, is the confidence of the rule 2->1 must also be 100%:  
☐ True.  
☒ False.
10. The K-means clustering algorithm will automatically find the best value of k (number of cluster):  
☐ True.  
☒ False.
11. Given the following three document out of 150 documents, and a table of  $df$  of terms of the three documents  
Doc 1="computer programming of computer university", Doc 2="programming in university",  
Doc 3="university of programming"

