Questions and Answers

1. By creating an index the data will increase in term of size (1pt):
   ☐ True.
   ☐ False.

2. Unstructured data means (1pt):
   ☐ Has clear semantic forms.
   ☐ Has not clear semantic forms.
   ☐ None of the above.

3. Information retrieval goal is (1pt):
   ☐ 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) (1pt):
   ☐ True.
   ☐ False.

5. Does best match IR model type, result a ranked list of documents (best result is the first) (1pt):
   ☐ True
   ☐ False

6. Exact matching may retrieve too few or too many documents (1pt):
   ☐ True
   ☐ False

7. In what situation rare terms have the same value of frequent terms (1pt):
   ☐ 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 (1pt):
   ☐ 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% (1.5pt):
   ☐ True.
   ☐ False.

10. The K-means clustering algorithm will automatically find the best value of k (number of cluster) (1.5pt):
    ☐ 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”

Word	computer	programming	of	university	in
df	80	50	150	30	120
idf					

- Calculate the values of idf (2pt)?
- Assuming the following query =”computer programming”, and SMART notation LNN.NTN. What are similarities between query and the first document (Doc 1) (2pt)?

12. Given the following 4 transaction:

TID	ITEMS
1	K, A, D, B
2	D, A, C, E, B
3	C, A, B, E
4	B, A, D

- Assuming min_sup = 60% and the confidence is 80%, extract all association rules (3pt)?

13. Given the following data set, draw a decision tree without do any calculation (2pt)?

A	B	C	Y
F	F	F	F
T	F	T	T
T	T	F	T
T	T	T	F