Birzeit University
Department of Electrical and Computer Engineering
Faculty of Engineering and Technology
ENCS539: NLP –IR   Quiz1 , September 30 , 2021

Student Name: ____AA_____   Student Number: _____25/25_____

Question 1: Consider the following small index (3 terms, 7 documents) with postings in the following format:
```
Word Type:DocumentFrequency:doc#; doc#;…. doc#.
T1:Black:      Df1:  1;    2;  3; 4.
T1:Sheep:      Df2:  3;    4;  7.
T1:Fish:       Df3:  1;  2;  3; 5.
```

1. Please give values for document frequencies for terms T1,T2,T3: 3/25
Answer: Df1=   4   Df2=   3   Df3   4

1.5. tf-idf for Black in Document1 $(tf_{1,1}\text{-}idf_1)$ –no normalization, no log- = ____**1x7/4=7/4**____2/25

2. Reconstruct the full **binary** Incidence Matrix (by filling 0 or 1): Use this for items 3 trough 8. 6/25

	Doc1	Doc2	Doc3	Doc4	Doc5	Doc6	Doc7
Black	1	1	1	1	0	0	0
Sheep	0	0	1	1	0	0	1
Fish	1	1	1	0	1	0	0

3. What is the **Cosine Similarity** between document1 and the query (Black, Fish, Sheep). 3/25
D1=(1,0,1); Q=(1,1,1)
Sim(D1,Q)= (1x1+1x0+1x1)/SQRT(2x3)=2/sqrt(6)=2/2.4=.81

4. What is the **Cosine Similarity** between document1 and document3. 3/25

D1=(1,0,1); D3=(1,1,1)
Sim(D1,Q)= (1x1+1x0+1x1)/SQRT(2x3)=2/sqrt(6)=2/2.4=.81
=
5- The Document vector for **Sheep** is: _0011001_____   and for **Fish** is:_1110100___ 2/25

6- The Term vector for **Doc1** is: __101_____   and for **Doc4** is:__110___ and for **Doc5** is:__001 2/25

7- The query **Black** AND **Sheep** has the answer set (list all matching documents): 2/25
{D3,D4}

8- The query **Black** AND **Fish'** (not **Fish**) has the answer set (list all matching documents): 2/25
{D4}