Exercise 1.

Document	Content
D_1	Big cats are nice and funny
D_2	Small dogs are better than big dog
D_3	Small cats are afraid of small dogs
D_4	Big cats are not afraid of small dogs
D_5	funny cats are not afraid of small dogs.

Step 1 :- Stemming & Stop word Removal
Step 2 :- Formation of Term x Document matrix
Step 3 :- check user query is in DNF form . If not convert to DNF
Step 4 :- Find similarity b/w query & each document by applying similarity function.

Step 1 :-

After Stemming & Stop word Removal { Big , Cat , nice , funny , Small , dog , better , afraid.

Step 2 :-

Keywords \ Documents	D1	D2	D3	D4	D5
big (K_1)	1	1	0	1	0
Cat (K_2)	1		1	1	1
nice (K_3)	1				
funny (K_4)	1				1
Small (K_5)		1	1	1	1
dog (K_6)		1	1	1	1
better (K_7)		1			
afraid (K_8)			1	1	1

Step 3 :-

$q = Funny \wedge DOg$ → $q_{DNF} = funny \wedge DOg$. $q_{CC} = funny \wedge DOg$

already in DNF

q_{cc} = funny $\wedge$ Dog

	K_4	K_6
D_1	1	0
D_2	0	1
D_3	0	1
D_4	0	1
D_5	1	1

Query is $\wedge$ Operator.

$Sim(d_j, q) = 1$ if $\exists q_{cc} | (q_{cc} \in \vec{Q_{dnf}}) \wedge (\forall k_i, g_i(\vec{d_j}) = g_i(q_{cc}))$

Two keywords first is K_4 & second is K_6

for query, $g_i(q_{cc}) = (1,1)$ as they are present

Only for D_5, $g_i(D_5) = (1,1)$ as they are present

Document D_5 is Retrieved

(2) q = funny $\vee$ Dog

$\vec{Q_{dnf}}$ = funny $\vee$ Dog q_{cc} = funny $\vee$ Dog

Document D_1, D_2, D_3, D_4, D_5 Retrieved

(3) q = Big $\wedge$ Dog $\wedge \neg$ funny

$\vec{Q_{dnf}}$ = Big $\wedge$ DOg $\wedge \neg$ funny

q_{cc} = Big $\wedge$ Dog $\wedge \neg$ funny.

q_{cc} = Big $\wedge$ Dog $\wedge \neg$ funny

	K_1	K_6	$\neg K_4$
D_1	1	0	1
D_2	1	1	0
D_3	0	1	0
D_4	1	1	0
D_5	0	1	1

Three keywords: K_1, K_6 & $\neg K_4$

for query $g_i(q_{cc}) = 1$ for (1,1,0)

Only for D2 & D4 matches

Document D2 & D4 Retrieved

Case 2.

Document	Content
D_1	Computer Information Retrival
D_2	Computer Retrival
D_3	Information
D_4	Computer Information

Step 1 :- Computer, Information, Retrival

Step 2 :-

	D_1	D_2	D_3	D_4
Computer (k_1)	1	1	0	1
Information (k_2)	1	0	1	1
Retrival (k_3)	1	1	0	0

q = information $\wedge$ Retrival ← D1 retrived.

q = information $\vee$ Retrival. ← All document Retrived

CNF

q = information $\wedge$ (Computer $\vee$ retrival)

q = (information $\wedge$ Computer) $\vee$ (information $\wedge$ Retrival)

↑ D1, D4

Step 3 : Check the query is in DNF form. If not convert to DNF.

$q = Cat \wedge (Nice \vee Afraid)$ ← CNF

$q = (Cat \wedge Nice) \vee (Cat \wedge Afraid)$ ← DNF

$\vec{q}_{dnf} = (Cat \wedge Nice) \vee (Cat \wedge Afraid)$

q_{cc} q_{cc}

Step 4 :

$\vec{q}_{dnf} = (Cat \wedge Nice) \vee (Cat \wedge Afraid)$

q_{cc} q_{cc}

$q_{cc} = (Cat \wedge Nia)$

	k_2	k_3
D_1	1	1
D_2	0	0
D_3	1	0
D_4	1	0
D_5	1	0

$q_{cc} = (Cat \wedge Afraid)$

	k_2	k_3
D_1	1	0
D_2	0	0
D_3	1	1
D_4	1	1
D_5	1	1

D_1, D_3, D_4, D_5 Retrieved