

Ejercicio 1 - Q1

TF-IDF (N=38, TF=1)

Terminos	ni	IDF = log(N/ni)	TF-IDF
72	1	5.25	5.25
117	1	5.25	5.25
191	5	2.93	2.93

Query 1: (72, 117, 191)

Tabla de IDF (Documentos)

Doc \ Term	72	117	191
36	5.25	0	0
17	0	5.25	0
5	0	0	2.93
15	0	0	2.93
25	0	0	2.93
29	0	0	2.93
32	0	0	2.93

Ranking de documentos

Doc	Rank	¿Info Needs?
36	0.66	SI
17	0.66	NO
5	0.37	SI
15	0.37	SI
25	0.37	SI
29	0.37	NO
32	0.37	NO

Tabla de IDF (Query)

Query \ Term	72	117	191
1	5.25	5.25	2.93

Podemos ver que se recuperaron los 4 documentos relevantes.

Documentos relevantes

Doc	¿Se recuperaron?
5	SI
15	SI
25	SI
36	SI

Ejercicio 1 - Q1 (Reescrita)

TF-IDF (N=38, TF=1)

Terminos	ni	IDF = log(N/ni)	TF-IDF
72	1	5.25	5.25
116	1	5.25	5.25
191	5	2.93	2.93

Tabla de IDF (Documentos)

Doc \ Term	72	116	191
36	5.25	0	0
5	0	5.25	2.93
15	0	0	2.93
25	0	0	2.93
29	0	0	2.93
32	0	0	2.93

Tabla de IDF (Query)

Query \ Term	72	116	191
1	5.25	5.25	2.93

Documentos relevantes

Doc	¿Se recuperaron?
5	SI
15	SI
25	SI
36	SI

Query 1: (72, 116, 191)

Ranking de documentos

Doc	Rank	¿Info Needs?
5	0.75	SI
36	0.66	SI
15	0.37	SI
25	0.37	SI
29	0.37	NO
32	0.37	NO

Observación:

Podemos ver que con la nueva query pudimos devolver los 4 documentos relevantes en las primeros posiciones

Ejercicio 1 - Q2

TF-IDF (N=38, TF=1)

Terminos	ni	IDF = log(N/ni)	TF-IDF
147	4	3.25	3.25
195	9	2.08	2.08
196	8	2.25	2.25

Query 2: (147, 195, 196)

Tabla de IDF (Documentos)

Doc \ Term	147	195	196
8	3.25	0	0
10	3.25	0	0
18	3.25	0	0
32	3.25	0	0
1	0	2.08	0
4	0	2.08	2.25
6	0	2.08	0
9	0	2.08	2.25
19	0	2.08	2.25
21	0	2.08	2.25
29	0	2.08	0
37	0	2.08	2.25
38	0	2.08	2.25
14	0	0	2.25
36	0	0	2.25

Ranking de documentos

Doc	Rank	¿Info Needs?
8	0.73	NO
10	0.73	NO
18	0.73	NO
32	0.73	NO
4	0.69	SI
9	0.69	SI
19	0.69	SI
21	0.69	SI
37	0.69	SI
38	0.69	SI
14	0.5	SI
36	0.5	SI
1	0.47	SI
6	0.47	SI
29	0.47	SI

Tabla de IDF (Query)

Query \ Term	147	195	196
2	3.25	2.08	2.25

Documentos relevantes

Doc	¿Se recuperaron?
1	SI
4	SI
6	SI
9	SI
14	SI
17	NO
19	SI
21	SI
23	NO
29	SI
36	SI
37	SI
38	SI

Ejercicio 1 - Q2 (Reescrita)

TF-IDF (N=38, TF=1)

Terminos	ni	IDF = $\log(N/n_i)$	TF-IDF
195	9	2.08	2.08
196	8	2.25	2.25

Tabla de IDF (Documentos)

Doc \ Term	195	196
1	2.08	0
4	2.08	2.25
6	2.08	0
9	2.08	2.25
19	2.08	2.25
21	2.08	2.25
29	2.08	0
37	2.08	2.25
38	2.08	2.25
14	0	2.25
36	0	2.25

Tabla de IDF (Query)

Query \ Term	195	196
2	2.08	2.25

Documentos relevantes

Doc	¿Se recuperaron?
1	SI
4	SI
6	SI
9	SI
14	SI
17	NO
19	SI
21	SI
23	NO
29	SI
36	SI
37	SI
38	SI

Query 2: (195, 196)

Ranking de documentos

Doc	Rank	¿Info Needs?
4	1	SI
9	1	SI
19	1	SI
21	1	SI
37	1	SI
38	1	SI
14	0.73	SI
36	0.73	SI
1	0.68	SI
6	0.68	SI
29	0.68	SI

Observación:

Podemos ver que con la nueva query pudimos devolver 11 de los 13 documentos relevantes, además de que no se devolvieron documentos no relevantes. En conclusion podemos ver una mejora sustancial con respecto a la Q2 original. Además destaco que solo tuvimos que eliminar un termino de la query para poder mejorar su precisión.

Ejercicio 1 - Q3

TF-IDF (N=38, TF=1)

Terminos	ni	IDF = log(N/ni)	TF-IDF
55	3	3.66	3.66
56	1	5.25	5.25
141	6	2.66	2.66
142	0	0	0
147	4	3.25	3.25

Query 3: (55, 56, 141, 142, 147)

Tabla de IDF (Documentos)

Doc \ Term	55	56	141	142	147
14	3.66	0	0	0	0
19	3.66	0	0	0	0
32	3.66	0	0	0	3.25
22	0	5.25	2.66	0	0
2	0	0	2.66	0	0
11	0	0	2.66	0	0
20	0	0	2.66	0	0
33	0	0	2.66	0	0
34	0	0	2.66	0	0
8	0	0	0	0	3.25
10	0	0	0	0	3.25
18	0	0	0	0	3.25

Ranking de documentos

Doc	Rank	¿Info Needs?
14	0.48	NO
19	0.48	NO
32	0.64	NO
22	0.77	SI
2	0.35	SI
11	0.35	SI
20	0.35	SI
33	0.35	SI
34	0.35	SI
8	0.42	NO
10	0.42	NO
18	0.42	NO

Tabla de IDF (Query)

Query \ Term	55	56	141	142	147
3	3.66	5.25	2.66	0	3.25

Documentos relevantes

Doc	¿Se recuperaron?
2	SI
3	NO
11	SI
12	NO
13	NO
16	NO
20	SI
22	SI
24	NO
26	NO
28	NO
33	SI
34	SI

Ejercicio 1 - Q3 (Reescrita)

TF-IDF (N=38, TF=1)

Terminos	ni	IDF = log(N/ni)	TF-IDF
56	1	5.25	5.25
141	6	2.66	2.66
144	3	3.66	3.66
145	1	5.25	5.25

Query 3: (55, 141, 144, 145)

Tabla de IDF (Documentos)

Doc \ Term	56	141	144	145
22	5.25	2.66	0	0
2	0	2.66	0	0
11	0	2.66	0	0
20	0	2.66	0	0
33	0	2.66	3.66	0
34	0	2.66	0	0
3	0	0	3.66	0
28	0	0	3.66	0
24	0	0	0	5.25

Ranking de documentos

Doc	Rank	¿Info Needs?
22	0.68	SI
24	0.6	SI
33	0.52	SI
3	0.42	SI
28	0.42	SI
2	0.31	SI
11	0.31	SI
20	0.31	SI
34	0.31	SI

Tabla de IDF (Query)

Query \ Term	56	141	144	145
3	5.25	2.66	3.66	5.25

Observación:

Podemos ver que con la nueva query pudimos devolver 9 de los 13 documentos relevantes, además de que no se devolvieron documentos no relevantes.

Documentos relevantes

Doc	¿Se recuperaron?
2	SI
3	SI
11	SI
12	NO
13	NO
16	NO
20	SI
22	SI
24	SI
26	NO
28	SI
33	SI
34	SI

Ejercicio 1 - Q4

TF-IDF (N=38, TF=1)

Terminos	ni	IDF = log(N/ni)	TF-IDF
147	4	3.25	3.25
179	5	2.93	2.93
180	0	0	0

Tabla de IDF (Documentos)

Doc \ Term	147	179	180
8	3.25	0	0
10	3.25	0	0
18	3.25	0	0
32	3.25	0	0
7	0	2.93	0
13	0	2.93	0
27	0	2.93	0
30	0	2.93	0
35	0	2.93	0

Tabla de IDF (Query)

Query \ Term	147	179	180
4	3.25	2.93	0

Documentos relevantes

Doc	¿Se recuperaron?
7	SI
27	SI
30	SI
31	NO
35	SI

Query 4: (147, 179, 180)

Ranking de documentos

Doc	Rank	¿Info Needs?
8	0.74	NO
10	0.74	NO
18	0.74	NO
32	0.74	NO
7	0.67	SI
13	0.67	NO
27	0.67	SI
30	0.67	SI
35	0.67	SI

Ejercicio 1 - Q4

TF-IDF (N=38, TF=1)

Terminos	ni	IDF = log(N/ni)	TF-IDF
179	5	2.93	2.93
54	2	4.25	4.25

Tabla de IDF (Documentos)

Doc \ Term	179	54
31	0	4.25
7	2.93	0
13	2.93	0
27	2.93	0
30	2.93	4.25
35	2.93	0

Tabla de IDF (Query)

Query \ Term	179	54
4	2.93	4.25

Documentos relevantes

Doc	¿Se recuperaron?
7	SI
27	SI
30	SI
31	SI
35	SI

Query 4: (179, 54)

Ranking de documentos

Doc	Rank	¿Info Needs?
31	0.82	SI
7	0.57	SI
13	0.57	NO
27	0.57	SI
30	1	SI
35	0.57	SI

Observación:

Podemos ver que con la nueva query pudimos devolver todos documentos relevantes, ademas de que se mejoró el ranking de los primeros 3 resultados los cuales en la query original eran documentos irrelevantes.

Ejercicio 1 - Q5

TF-IDF (N=38, TF=1)

Terminos	ni	IDF = log(N/ni)	TF-IDF
147	4	3.25	3.25
182	4	3.25	3.25
184	4	3.25	3.25

Tabla de IDF (Documentos)

Doc \ Term	147	182	184
8	3.25	3.25	3.25
10	3.25	3.25	3.25
18	3.25	3.25	3.25
32	3.25	3.25	3.25

Tabla de IDF (Query)

Query \ Term	147	182	184
5	3.25	3.25	3.25

Documentos relevantes

Doc	¿Se recuperaron?
8	SI
10	SI
18	SI
32	SI

Query 5: (147, 182, 184)

Ranking de documentos

Doc	Rank	¿Info Needs?
8	1	SI
10	1	SI
18	1	SI
32	1	SI

Observación:

Podemos ver que con la nueva query original ya pudimos devolver todos los documentos relevantes.