

## Ranking Retrieval

Profesor Heider Sanchez

FABRIZIO FRANCO AMAYO (201710466)

**P1. Llene el siguiente cuadro, analice los resultados y de una explicación de dicho comportamiento.**

	(Q, Doc1)	(Q, Doc2)
Score1	3.93	3.73
Score2	0.69	0.78
Explicación	Para el Score 1, se encuentra que Doc 1 y QUE son más similares porque se toma en cuenta tanto la frecuencia como la cantidad de términos que aparecen. Para el score 2, Doc 2 y Q son más parecidos porque los términos de intersección son más relevantes que en el otro documento.	

**P2. Llene el siguiente cuadro, analice los resultados y de una explicación de dicho comportamiento.**

	(Q, Doc1)	(Q, Doc2)
DotProduct	35.31	27.45
cosine	0.52	0.62
Explicación	El score 1, de Doc1 es mayor por la frecuencia de los términos si se compara con la frecuencia de los términos en Doc 2, respecto al contenido del query en ambos casos. Para el score 2, Doc 2 tiene un mayor puntaje porque los términos de intersección son más importantes que aquellos presentes en Doc 1.	

**P3. Esquematiza el funcionamiento del índice invertido para dar soporte al modelo de ranked retrieval.**

W1	idf	DOC 1, TF(DOC1)	DOC 4, TF(DOC4)	...
W2	idf	DOC 2, TF(DOC2)	DOC 10, TF(10)	...
W3	idf	DOC 3, TF(DOC3)	DOC 11, TF(DOC11)	...
⋮				
⋮				
⋮				

DOC 1	DOC1
DOC 2	DOC2
DOC 3	DOC3
DOC 4	DOC4

Para construir el índice se itera los documentos linealmente por lo que se va incrementando un contador por término para al finalizar hallar el idf, conforme se itera cada documento, ~~por bloque~~ al agregar palabras al índice invertido, se escribe los documentos por término en memoria principal, una vez se haya terminado de analizar un documento se calcula el tf de cada palabra en ese documento.