

1) Data una collezione di documenti e le relative occorrenze dei termini:

$d_1 = (T1:7, T2:5, T5:1)$

$d_2 = (T1:4, T5:3)$

$d_3 = (T1:2, T3:2, T4:1)$

$d_4 = (T3:2)$

a) Costruire l'indice invertito della collezione

b) Calcolare il ranking dei documenti rispetto alla query $q = (T1:1 \text{ AND } T5:2)$ utilizzando la similarità del prodotto interno e utilizzando le occorrenze dei termini per la rappresentazione dei documenti.

a) Costruzione indice invertito

$T1 \rightarrow [D1 | D2 | D3]$

$T2 \rightarrow [D1]$

$T3 \rightarrow [D3 | D4]$

$T4 \rightarrow [D3]$

$T5 \rightarrow [D1 | D2]$

b) Per risolvere Q seguiamo queste procedure:
Recuperare la posting list di $T1$

- Fare l'intersezione tra le 2 posting list $T5$

- RISULTATO $D1, D2$

Calcolo la SIM. del prodotto interno tra

$$SIM(Q, D1) = 7 \times 1 + 5 \times 0 + 0 \times 0 + 0 \times 0 + 1 \times 2 = 9$$

$$SIM(Q, D2) = 4 \times 1 + 0 \times 0 + 0 \times 0 + 0 \times 0 + 3 \times 2 = 10$$

$$\vec{D1} = \begin{pmatrix} T1 & T2 & T3 & T4 & T5 \\ 7 & 5 & 0 & 0 & 1 \end{pmatrix}$$

$$\vec{D2} = \begin{pmatrix} 4 & 0 & 0 & 0 & 3 \end{pmatrix}$$

$$\vec{Q} = \begin{pmatrix} 1 & 1 & 0 & 0 & 2 \end{pmatrix}$$

$$\vec{Q} = (1 \quad 0 \quad 0 \quad 0 \quad 2)$$

Calcoliamo la SIM del COSENDO

$$\cos(\vec{D}_1, \vec{Q}) = \frac{\vec{D}_1 \cdot \vec{Q}}{\|\vec{D}_1\| \|\vec{Q}\|} = \frac{9}{\sqrt{75} \sqrt{5}} = 0.465$$

$$\cos(\vec{D}_2, \vec{Q}) = \frac{\vec{D}_2 \cdot \vec{Q}}{\|\vec{D}_2\| \|\vec{Q}\|} = \frac{10}{5\sqrt{5}} = 0.894$$

$$\|\vec{D}_1\| = \sqrt{7^2 + 5^2 + 1^2} = \sqrt{75}$$

$$\|\vec{D}_2\| = \sqrt{4^2 + 3^2} = \sqrt{25} = 5$$

$$\|\vec{Q}\| = \sqrt{1^2 + 2^2} = \sqrt{5}$$

$D1 = \langle \text{AUTOMOBILE IN CODA PER VIA DEL TRAFFICO, BLOCCO TRAFFICO} \rangle$

$D2 = \langle \text{SEMAFORO ROSSO E TRAFFICO INTENSO} \rangle$

$Q = \langle \text{TRAFFICO INTENSO} \rangle$

Calcolare le SIM tra Q e $D1$ e $D2$

Calcoliamo le BAG OF WORDS

$D1 = (\text{AUTOMOBILE:1, CODA:1, TRAFFICO:2, BLOCCO:1})$

$D2 = (\text{SEMAFORO:1, Rosso:1, TRAFFICO:1, INTENSO:1})$

$Q = (\text{TRAFFICO:1, INTENSO:1})$

	AUT	CODA	TRAF	Blocco	SEMAF	Rosso	INTENSO
$D1$	(1	1	2	1	0	0	0)
$D2$	(0	0	1	0	1	1	1)
Q	(0	0	1	0	0	0	1)

$$\text{COSIM}(D1, Q) = \frac{1 \times 0 + 1 \times 0 + 2 \times 1 + 1 \times 0 + 0 \times 0 + 0 \times 0 + 0 \times 1}{\|D1\| \|Q\|} = \frac{2}{\sqrt{7} \sqrt{2}} = 0.534$$

$$\text{COSIM}(D2, Q) = \frac{0 \times 0 + 0 \times 0 + 1 \times 1 + 0 \times 0 + 1 \times 0 + 1 \times 1 + 1 \times 1}{\|D2\| \|Q\|} = \frac{3}{2 \sqrt{2}} = 0.707$$

$$\|D1\| = \sqrt{1^2 + 1^2 + 2^2 + 1^2} = \sqrt{7}$$

$$\|D2\| = \sqrt{1^2 + 1^2 + 1^2 + 1^2} = \sqrt{4} = 2$$

$$\|Q\| = \sqrt{1^2 + 1^2} = \sqrt{2}$$