

- 1) Sia q una query i cui documenti rilevanti nella collezione sono 5. Siano S_1 ed S_2 due sistemi che riportano i seguenti primi 10 risultati in risposta alla query q . R indica che un documento è rilevante, N indica che il documento è non rilevante (il risultato più a sinistra è il top della lista).

S_1 : RNRNRNRNR

S_2 : NRNRNRNRNR

Calcolare l'accuratezza dei due sistemi per la query q , utilizzando le seguenti metriche, fornendone anche una breve descrizione:

- a) $P@1$, $P@5$, $P@10$, R-precision

(PUNTI 4)

- b) Average Precision

(PUNTI 4)

- c) **Facoltativo:** Riportare la curva di precisione-richiamo per la query q per il sistema S_1

(PUNTI 3)

$$P@1(S_1) = 1 \quad P@5(S_1) = 2/5 \quad P@10(S_1) = 5/10$$

$$R-p(S_1) = P@5 = 2/5$$

$$P@1(S_2) = 0 \quad P@5(S_2) = 2/5 \quad P@10(S_2) = 4/10 \quad R-p(S_2) = P@5 = 2/5$$

$$AvP(S_1) = \frac{1 + 2/3 + 3/6 + 4/9 + 5/10}{5} = 0.62$$

$$AvP(S_2) = \frac{1/2 + 2/5 + 3/6 + 4/7 + 0}{5} = 0.4$$

S_1 : X O X O O X O O X X

$P=1$ $P=2/3=0.66$ $P=3/6=0.5$ $P=4/9=0.44$ $P=5/10=0.5$
 $R=1/5=0.2$ $R=2/5=0.4$ $R=3/5=0.6$ $R=4/5=0.8$ $R=5/5=1$

$$P(\pi_j) = \max_{\pi_j \geq \pi_i} R(\pi_i)$$

P	R
1	0
1	0.1
1	0.2
0.666	0.3
0.666	0.4
0.5	0.5
0.5	0.6
0.5	0.7
0.5	0.8
0.5	0.9
0.5	1

2) Sia q una query che ha 8 documenti rilevanti nella collezione. Supponiamo che un algoritmo di ritrovamento applicato a q riporti il seguente ranking R_q : D3 D1 D5 D7 D9

Supponiamo che D1 e D7 siano documenti rilevanti per q .

Calcolare *Precision*, *Recall*, *F1*, *R-precision* ed *Average Precision* per q , fornendo anche una descrizione formale delle metriche

(PUNTI 8)

D3	D1	D5	D7	D9
0	X	0	X	0

$$P = 2/5$$

$$R = 2/8$$

$$F1 = \frac{2 \times R \times P}{P + R} = \frac{2 \times 2/5 \times 2/8}{2/5 + 2/8}$$

$$R-p = P@8 = \frac{2}{8}$$

$$AP = \frac{1/2 + 2/4}{8}$$