1. Si consideri una collezione costituita da 700 documenti e due query *q1* e *q2*

Si supponga di sapere che:

* q1 ha come risultato un insieme di 70 documenti, di cui 38 rilevanti;
* q2 ha come risultato un insieme di 50 documenti, di cui 25 rilevanti;

Quale delle due query conviene formulare? Perché?

Se si sapesse anche che per ognuna delle query i documenti rilevanti nella collezione sono 90, la risposta potrebbe cambiare? Perché?

1. Descrivere i problemi di *polisemia* e *sinonimia* e spiegare che influenza hanno sulle metriche di precisione e richiamo.
2. Descrivere l’ontologia linguistica WordNet ed indicare come è organizzata e rappresentata la conoscenza lessicale al suo interno.
3. Sia dato il documento *d* contenente i seguenti termini con relative occorrenze:

*a) esame: 4*

*b) gestione: 2*

*c) conoscenza: 1*

Si assuma di avere una collezione di 5000 documenti in cui:

*a) il numero di documenti contenenti il termine “esame” è 800;*

*b) il numero di documenti contenenti il termine “gestione” è 400;*

*c) il numero di documenti contenenti il termine “conoscenza” è 10.*

Calcolare per ognuno dei tre termini in *d* il corrispondente valore tf-idf.

1. Descrivere brevemente i principali moduli di un sistema di *Information Retrieval*.
2. Descrivere brevemente le principali caratteristiche dei modelli di ritrovamento *booleano* e a *spazio vettoriale*, evidenziando le principali differenze.