Multimedia testo - Esercizio

Si consideri il seguente testo:

> Una base di dati é una collezione di insiemi omogenei di dati memorizzata su memoria permanente ed utilizzabile tramite un DBMS (Data Base Management System, sistema di gestione di basi di dati).

contenuto in una base di documenti.

1. Identificare un insieme di termini indice per tale testo.
2. Supponendo una base di documenti con 20 termini indice, contenente 200 documenti, mostrare la rappresentazione del documento nel modello booleano e in quello vettoriale, e discutere in entrambi i modelli la formulazione e la valutazione dell'interrogazione in cui si ricercano i termini "base" e "dato" (si supponga che dei 200 documenti 10 contengano il termine "base" e 50 il termine "dato").
3. Supponendo il testo memorizzato nella colonna C di una tabella TAB, su cui sia definito un indice text, formulare in Oracle le seguenti interrogazioni:
   (a) trovare i documenti che contengono sia base che dati, ordinati in base al ranking;
   (b) trovare i documenti che contengono sia base che dati, privilegiando nel ranking i documenti che contengono entrambi i termini;
   (c) trovare i documenti che contengono sia base che dati, privilegiando nel ranking i documenti che contengono base;
   (d) trovare i documenti che contengono base e dati, vicini il piú possibile e in tale ordine;
   (e) trovare i documenti che contengono base e dati, o termini che sono variazioni della stessa radice;
   (f) trovare i documenti che contengono base e dati, o i loro equivalenti termini inglesi (si supponga caricato un Thesaurus contenente la traduzione di tali termini).