Basi di Dati: Evoluzioni Recenti

Quarto Compitino - Parte Testo

31 Maggio 2001

1. Si consideri il seguente testo:

   I dati sono trasferiti tra il disco e la memoria principale in unità chiamate blocchi. Un blocco è una sequenza di byte contigui memorizzati in una stessa traccia dello stesso cilindro. Il tempo di trasferimento di un blocco è il tempo impiegato dalla testina per trasferire un blocco in memoria principale, una volta posizionata all'inizio del blocco.

   (a) (Indicativamente, punti 5) Illustrare su tale testo le varie fasi di elaborazione del testo, discutendo come il testo viene modificato da ciascuna fase.

   (b) (Indicativamente, punti 9) Sia d_1 il documento al punto precedente e

   d_2 = Una base di dati è una collezione di dati tra loro correlati, memorizzati su memoria secondaria e gestita da un sistema di gestione di basi di dati.

   d_3 = In un sistema bancario i dati comprendono informazioni relative ai clienti, ai conti correnti con i rispettivi saldi, e le operazioni effettuate su tali conti.

   d_3 = Il blocco contenente la porzione di file di log relativa ad una certa modifica, deve essere forzato su memoria stabile prima di copiare su memoria secondaria il blocco contenente il dato cui è relativa la modifica.

   d_5 = Una base di dati temporale è una base di dati in cui viene memorizzata l'evoluzione dei dati nel tempo.

   Fissati inoltre i seguenti termini indice: dato, memoria, tempo, sistema, blocco, conto

      i. mostrare:
         A. la struttura di inverted index per la base di documenti,
         B. la rappresentazione dei documenti nel modello booleano,
         C. la rappresentazione dei documenti nel modello vettoriale,
      ii. rappresentare e valutare nel modello booleano l'interrogazione per ritrovare i documenti che trattano sia di dati, che di memoria, che di (tempo) o (blocchi ma non sistemi))
      iii. rappressentare nel modello vettoriale l'interrogazione per ritrovare i documenti che trattano di conti e valutarla su uno dei documenti

   (c) (Indicativamente, punti 6) Supponendo i documenti memorizzati nella colonna C di una tabella TAB, su cui sia definito un indice text, formulare in Oracle le seguenti interrogazioni:

      i. trovare i documenti che contengono nella stessa frase sia memoria che tempo o variazioni delle rispettive radici;
      ii. trovare i documenti che contengono sia sistema che tempo, eliminando i documenti in cui il ranking di tempo è inferiore a 10 e privilegiando nel ranking i documenti che contengono più occorrenze di sistema;
      iii. trovare i documenti che contengono sistema e/o tempo, privilegiando nel ranking i documenti che contengono entrambi i termini, ordinando i documenti restituiti in base al ranking;
      iv. trovare i documenti che contengono (tempo e non blocco) oppure (tempo e preferibilmente non sistema);
      v. trovare i documenti che contengono blocco e memoria, vicini il piú possibile e separati da non più di 4 parole ma non necessariamente in tale ordine;
      vi. trovare i documenti che contengono un temine di cui DBMS è un termine più ristretto di tipo istanza (si supponga caricato un Thesaurus contenente la corrispondente relazione tra tali termini).

```
select id
from texttab
where contains(textcol, ´query´) > 0

select id, score(Num)
from texttab
where contains(textcol, `query´, Num) > 0
order by score(Num) desc
```

- **Operatori booleani**
  - AND, &
  - OR, |
  - NOT, ~
- **Operatori di score**
  - $\star$ (weight/peso — secondo argomento è numero tra .1 e 10)
  - $>$ (threshold/soglia)
  - - o MINUS (_ meglio se non _)
  - , o ACCUM (_ o _, meglio se entrambi)
- **Operatori di espansione dei termini**
  - % (wildcard)
  - ? (espansione fuzzy)
  - $ (stem)
  - ! (soundex)
  - =, EQUIV
- **Operatori di prossimità**
  - ;
  - NEAR((lista di parole separate da ,), MaxDist, True/False)
- **Altri operatori**
  - WITHIN section
  - ABOUT()
  - operatori di Thesaurus
    - SYN(term, [,thesname]) sinonimi
    - PT(term, [,thesname]) preferred term
    - BT/BTP/BTG/BTI/NT/NTP/NTG/NTI(term, [,level [,thesname]]) broader/narrower term (G= generico, P = parte, I = instance)
    - TT(term, [,thesname]) top term
    - RT(term, [,thesname]) related term
    - TR(term, [,lang [,thesname]]) traduzione
    - TRSYN(term, [,lang [,thesname]]) sinonimi + traduzione