Algoritmi e Strutture Dati

Documentazione relativa al progetto di laboratorio

Studente

Alfonsi Andrea

Matr. 0000766102

Data di consegna: 12/12/2018

Ambiente: Windows, Visual Studio 2015

Indice

Strutture Dati	pag.2
Funzioni e Algoritmi	pag.3

Strutture dati utilizzate

Le strutture dati create per completare questo progetto sono le seguenti:

Struttura dati usata per rappresentare il valore di un attributo di un elemento del database.

Comprende un attributo 'val' che corrisponde al valore dell'elemento e un attributo 'next' che punta al successivo elemento della lista.

Struttura dati che rappresenta una 'riga' del database.

In particolare, ha un attributo 'row' che punta al primo elemento della riga e un attributo 'next' che punta alla riga successiva.

Semplice lista con una variabile 'data' di tipo intero e un puntatore all'elemento successivo. Durante lo svolgimento del progetto è stata particolarmente utile per memorizzare temporaneamente gli indici delle colonne richieste.

Particolare tipo di struttura dati usata per rappresenta il risultato di una query che utilizza il 'group by'.

Contiene una variabile 'row' contenente il nome della riga, una variabile 'count' che indica il numero di volte che l'elemento 'row' è presente nel database e un puntatore all'elemento successivo.

Funzioni e algoritmi utilizzati

-executeQuery:

Funzione che prende in input una stringa contenente la query da eseguire e che restituisce 'true' solo se l'esecuzione della query è andata a buon fine.

Come è possibile vedere la funzione usa uno switch e richiama la funzione 'queryType' che a sua volta determina il tipo di query da eseguire.

In base al risultato richiamerà poi le rispettive funzioni.

-queryType:

Come detto in precedenza questa funzione ha il compito di determinare il tipo di query richiesta.

Per farlo esegue una copia della variabile contenente la query, ne estrae la prima parola usando la funzione 'strtok' e resituisce quindi il tipo attraverso una variabile intera denominata 'type'.

-insert:

Funzione che inserisce una nuova riga all'interno di una tabella.

Cicla tutte le parole della query salvando le informazioni utili come il nome della tabella ed effettuando vari controlli.

```
while (words != NULL) {
         if (i == 2) tableName = strdup(words);
ᆸ
             char* tmp = strdup(words);
             tmp[strlen(tmp) - 1] = 0;
             ptr = fopen(strcat(tableName, ".txt"), "r");
             if (ptr == NULL) return 0;
             columns = (char*)malloc((17 + strlen(tableName) + strlen(tmp)) * sizeof(char*));
             fscanf(ptr, "%[^\n]", columns);
             fclose(ptr);
             columns = strtok(columns, " ");
             while (columns != NULL && j<4) {
                 if (j == 3) {
                     columns[strlen(columns) - 1] = 0;
                     if (strcmp(tmp, columns) != 0) return false;
                  else columns = strtok(NULL, " ");
```

In particolare controlla che le colonne corrispondano e che il numero di elementi inseriti sia lo stesso.

```
//se è giusto inserisco in coda al file

if (i == j) {
    ptr = fopen(tableName, "a");
    if (ptr == NULL) return 0;
    fore

frintf(ptr, row);
    fclose(ptr);
    return true;

else return false;
```

Se non ci sono errori, la funzione procede con l'inserire la nuova riga generata nel file selezionato e restituisce 1.

-save_ds:

Questa funzione ha lo scopo di salvare tutti gli elementi di una tabella in una struttura dati per poi restituirla.

```
DB * save_ds(char * table_name) {

FILE * ptr = fopen(strcat(table_name, ".txt"), "r");

if (ptr == NULL) return 0;

int line_size = 300;

char * line = calloc(line_size, sizeof(char*));

table_name[strlen(table_name) - 4] = 0;

DB * DB_head = NULL;

DB_head = calloc(1, sizeof(DB));

if (DB_head == NULL) return 0;

DB * DB_current = DB_head;

//salto la prima riga
fscanf(ptr, "%*[^\n]");

char * a = fgetc(ptr);
```

Nella prima parte della funzione viene aperto il file, viene saltata la prima riga e inizializzata la struttura dati che andrà a contenere i dati.

```
while (fgets(line, line size, ptr) != NULL) {
   char * current_line = strdup(line);
   current_line[strlen(current_line) - 2] = 0; //rimuovo gli ultimi due caratteri della riga
   current_line = strtok(current_line, " ");
current_line = strtok(NULL, " "); //salto la prima parola così da ottenere gli elementi da inserire nella struttura dati
   current_line = strtok(current_line, ","); //suddivido gli elementi singolarmente
   DB_current->row = calloc(1, sizeof(node_t));;
   if (DB_current->row == NULL) return 0;
   node_t * current_DB = DB_current->row;
    while (current_line != NULL) { //inserisco gli elementi nella struttura dati
       current_DB->val = current_line;
        current_DB->next = calloc(1, sizeof(node_t));
       current_DB = current_DB->next;
        current_line = strtok(NULL, ",");
   DB_current->next = calloc(1, sizeof(DB));
   DB_current = DB_current->next;
return DB head;
```

Nella seconda parte invece, viene ciclata ogni riga del file e per ognuna di esse se ne fa una copia che verrà poi ciclata a sua volta in modo da iterare tutti gli elementi di tale riga e da salvarli così nella struttura dati. Il costo sarà di conseguenza il numero di righe moltiplicato per il numero delle colonne.

Infine la struttura dati completa viene restituita.

-filter_columns:

Questa funzione è utilizzata quando è necessario lavorare solo con alcune delle colonne di una tabella.

Prende in input la tabella completa, le colonne da filtrare e il nome della tabella e in output restituisce la tabella filtrata.

```
pDB * filter_columns(DB * db_head, char * columns1, char * table name) {
     char * columns = strdup(columns1);
     char * words = strdup(columns);
     columns = strtok(columns, ",");
     node_t * head = malloc(sizeof(node_t));
     if (head == NULL) return 0;
     node_t * current = head;
     //ciclo ogni parola dentro columns e la inserisco nella lista
     while (columns != NULL) {
₫
         current->val = columns;
         columns = strtok(NULL, ",");
         if (columns != NULL) {
₿
             current->next = malloc(sizeof(node_t));
             current = current->next;
         else current->next = NULL;
```

Per prima cosa la funzione cicla le colonne e le inserisce in una lista.

```
//leggo dal file la prima riga e salvo gli indici
table_name = strcat(table_name, ".txt");

FILE * ptr = fopen(table_name, "r");

if (ptr == NULL) return 0;

columns = (char*)malloc((17 + strlen(table_name) + strlen(words)) * sizeof(char*));

fscanf(ptr, "%[^\n]", columns);

fclose(ptr);

columns += 11 + strlen(table_name);

table_name[strlen(table_name) - 4] = 0;

columns[strlen(columns) - 1] = 0;

//ho in head le colonne richieste dalla query, ho in columns le colonne della tabella
```

A questo punto il programma apre il file e legge la prima riga in modo da salvare tutte le colonne della tabella. Finito questo si avranno salvate in due strutture diverse sia le colonne richieste dalla query sia le colonne complete della tabella. Successivamente si andrà a creare una lista di tipo nodelnd che avrà lo scopo di contenere gli indici delle colonne richieste.

```
//while che salva gli indici delle colonne che mi interessano

while (current != NULL) {

char * dup_columns = strdup(columns);

dup_columns = strtok(dup_columns, ",");

cont = 0;

int changed = 0;

while (dup_columns != NULL && current != NULL) {

changed = 0;

if (strcmp(current->val, dup_columns) == 0) {

if (curr_q->data != -1) {

curr_q->next;

}

curr_q->data = cont;

changed = 1;

}

cont++;

if (changed == 1)

current = current->next;

dup_columns = strtok(NULL, ",");

}

153

}
```

Nell'immagine precedente è possibile vedere il ciclo che salva nella nuova lista gli indici delle colonne interessate.

```
while (db_current->row != NULL) {
         curr_q = q;
         node_t * old_node=db_current->row;
         node_t * new_node = new_db_current->row;
while (old_node != NULL && curr_q!=NULL) {
             if (curr_q->data == cont) {
                 new_node->val = old_node->val;
                 new_node->next = calloc(1, sizeof(node_t));
                 if (new_node->next == NULL) return 0;
                 new_node= new_node->next;
                 curr_q = curr_q->next;
                 cont = 0;
                 old_node = db_current->row;
             old_node = old_node->next;
             cont++;
         db_current = db_current->next;
         new_db_current->next = calloc(1, sizeof(DB));
         new_db_current = new_db_current->next;
         new_db_current->row = calloc(1, sizeof(node_t));
```

Infine si andrà a creare una nuova struttura dati di tipo 'DB' che conterrà la nuova tabella filtrata.

Il costo nel caso peggiore sarà pari al numero di righe nella tabella moltiplicato per il numero di colonne nella vecchia tabella mentre nel caso migliore sarà uguale al numero di righe per il numero di colonne nella nuova tabella e quindi varierà in base agli indici delle colonne nella nuova tabella.

Successivamente sarà restituita la tabella filtrata.

-operator_to_use:

Questa semplice funzione restituisce un numero che corrisponde all'operazione ricevuta in input ed è utilizzata quando è presente il filtro 'WHERE' nella query.

-get_columns:

Questa funzione legge la tabella ricevuta in input, ne salva la prima riga e la filtra in modo da poter restituire tutte le colonne di tale tabella.

-where:

Questa funzione applica il filtro where restituendo poi la tabella filtrata.

Per prima cosa la funzione trova l'indice della colonna su cui effettuare il controllo e ne salve l'indice in una variabile denominata 'index',

Successivamente andrà a ciclare tutte le righe della tabella e per ognuna di esse controllerà se il filtro è soddisfatto richiamando la funzione 'operator to use'.

```
int satisfied = 0;
白
          switch (operator_to_use(operation)) {
          case 1: if (strcmp(current_node->val, elem_2) == 0)
              satisfied = 1;
              break;
          case 2: if (strcmp(current_node->val ,elem_2)>0)
              satisfied = 1;
          case 3: if (strcmp(current node->val , elem 2)>=0)
              satisfied = 1;
          case 4: if (strcmp(current node->val , elem 2)<0)</pre>
              satisfied = 1;
              break;
          case 5: if (strcmp(current node->val , elem 2)<=0)</pre>
              satisfied = 1;
          default: return 0;
              break;
```

In caso il filtro sia soddisfatto si andrà ad aggiungere la riga corrente alla nuova tabella ed infine si restituirà.

```
//se il filtro è rispettato, aggiungo la riga corrente al nuovo db
if (satisfied == 1) {
    current_new_db->row = db_current->row;
    current_new_db->next = calloc(1, sizeof(DB));
    current_new_db = current_new_db->next;
}
db_current = db_current->next;
complete_db = complete_db->next;
}

return new_db;
}
```

-get_lval:

Questa funzione è utilizzata nel caso la query richieda un 'ORDER BY' in quanto l'algoritmo scelto per l'ordinamento è stato il 'quicksort' ed in particolare serve per restituire il valore di un particolare nodo della tabella.

Cicla le righe fino a trovare quella con l'indice richiesto e dopodiché scorre le colonne di tale riga fino a trovare quella desiderata andando poi a restituirla.

Il costo nel caso peggiore può essere il numero di righe sommato al numero di colonne mentre nel caso migliore è 1.

-swap:

Questa funzione è utilizzata durante l'ordinamento nel caso di una query con filtro 'ORDER BY' e serve a cambiare la posizione di due righe della tabella.

```
□void swap(DB * head, int i, int j) {
     DB * tmp = head;
     node_t * tmpival;
     node_t * tmpjval;
ൎ
     while (tmp && i) {
          tmp = tmp->next;
     tmpival = tmp->row;
     tmp = head;
₫
     while (tmp && j) {
          tmp = tmp->next;
      tmpjval = tmp->row;
      tmp->row = tmpival;
      tmp = head;
     while (tmp && i) {
          tmp = tmp->next;
      tmp->row = tmpjval;
```

La funzione comincia ciclando la tabella fino ad arrivare alla prima riga desiderata e la salva un una variabile 'tmpival'. Poi fa la stessa cosa con la seconda riga e infine ne cambia la posizione.

-quick_sort_list:

```
□DB * quick_sort_list(DB * head, int 1, int r, int index, char * order) {
       int i, j;
char * jval;
char * pivot;
П
           pivot = get_lval(head, 1, index)->val; //ricavo il valore per il pivot
            for (j = l + 1; j < r; j++) {
                jval = get_lval(head, j, index)->val;
if (strcmp(order, "ASC") == 0) {
 if (strcmp(jval, pivot) < 0 && jval != -1) {
                         swap(head, i, j);
                         i++;
 0-0
                else if (strcmp(order, "DESC") == 0) {
                     if (strcmp(jval, pivot) > 0 && jval != -1) {
                         swap(head, i, j);
            swap(head, i - 1, 1);
            quick_sort_list(head, l, i, index, order);
            quick_sort_list(head, i, r, index, order);
```

Questa funzione esegue l'ordinamento della tabella utilizzando un pivot e mettendo gli elementi più piccoli alla sua sinistra e quelli più grandi alla sua destra (o viceversa nel caso di ordinamento decrescente) facendo anche uso della funzione 'swap' per invertire la posizione delle righe.

Dopo aver ciclato tutti gli elementi ed aver ultimato l'ordinamento restituisce la tabella ordinata.

-order-by:

Per prima cosa, questa funzione trova l'indice della colonna in base alla quale effettuare l'ordinamento.

```
DB * order_by(DB * db_head, char * condition, char * order, char * table_name //trovo l'indice della colonna sul quale fare l'ordinamento int cont = 0, index = -1;

char * columns = get_columns(strdup(table_name));

columns = strtok(columns, ",");

while (columns != NULL) {

if (strcmp(columns, condition) == 0) {

index = cont;

break;

}

cont++;

columns = strtok(NULL, ",");
```

Successivamente conta il numero di elementi(righe) e poi lancia il *quicksort* passandogli tutte le variabili necessarie.

```
//conto il numero di elementi

B * tmp = db_head;
int n = 0;

while (tmp->row) {
    n++;
    tmp = tmp->next;

//uso il quick sort

db_head = quick_sort_list(db_head, 0, n, index, order);

return db_head;
```

Infine restituisce la tabella ordinata.

-group_by:

Questa funzione ha lo scopo di effettuare l'operazione di 'group by'. Per farlo comincia eliminando tutte le colonne che non gli interessano (attraverso la funzione 'filter columns') e creando la nuova struttura dati di tipo 'node group by'.

Successivamente cicla tutto la tabella e per ogni riga controlla se un elemento con lo stesso valore è gia presente nella nuova struttura dati e in tal caso ne aumenta il valore 'count'. Altrimenti aggiunge un nuovo nodo alla struttura dandogli valore 'count' uguale ad 1.

-print_query_gb:

Funzione con lo scopo di stampare nel file 'query_results' il risultato di una query che comprende l'operazione di 'group by'.

```
### Description of the print of the pr
```

Una volta aperto il file ci stampa la prima riga e poi cicla la struttura dati per tutte le righe seguenti; di conseguenza il costo è pari al numero di nodi nella struttura (che a sua volta equivale al numero di elementi distinti nella colonna della tabella sulla quale si è applicata la query di 'group by').

-print_query:

Questa funzione stampa il risultato di una query (tranne le query con il filtro 'group by').

Per farlo comincia aprendo il file e stampando la prima riga. Nel caso nella query vi fosse un 'SELECT *' le colonne devono essere ricavate attraverso la funzione 'get columns'.

Dopodiché cicla tutta la struttura dati stampando tutte le sue righe.

```
□void print_query(DB * db_head, char * query, char * table_name, char * columns) {

FILE * ptr = fopen("query_results.txt", "a");
      if (ptr == NULL) return 0;
      if (strcmp(columns, "*")==0) columns = get_columns(table_name);
      fprintf(ptr, "%s;\nTABLE %s COLUMNS %s;\n", query, table_name, columns);
      while (db_head!=NULL && db_head->row != NULL && db_head->row->val!=NULL) {
          fprintf(ptr, "ROW ");
          node_t * node = db_head->row;
          int first = 1;
          while (node->val != NULL) {
              if(first!=1)
                  fprintf(ptr, ",%s", node->val);
               else {
                  first = 0;
                   fprintf(ptr, "%s", node->val);
              node = node->next;
          fprintf(ptr, ";\n");
          db head = db head->next;
      fprintf(ptr, "\n");
      fclose(ptr);
```

-select:

La funzione select esegue le query di ricerca all'interno delle tabelle.

Per prima cosa cicla tutte le parole della query salvando varie informazioni come il nome della tabella, le colonne interessate dalla query ed eventuali filtri.

```
//ciclo tutte le parole nella query

while (query != NULL) {

if (i == 1) columns = strdup(query); //salvo le colonne richieste

else if (i == 3) table_name = strdup(query); //salvo il nome della tabella

if (i == 4) { //salvo il filtro e le variabili che mi serviranno

filter = strdup(query);

if (strcmp(filter, "WHERE") == 0) { ... }

else if (strcmp(query, "ORDER") == 0) { ... }

else if (strcmp(query, "GROUP") == 0) { ... }

i++;

query = strtok(NULL, " ");

494

}
```

A questo punto gestisco gli eventuali filtri richiamando le rispettive funzioni e inviandogli tutte le variabili necessarie.

```
//inserisco tutti gli elementi della tabella richiesta in una struttura dati

DB * db_head = save_ds(table_name);
node_group_by * db_head_groupby=NULL;
//gestisco filtri

switch (type) {
    case 1: db_head = where(db_head, elem_1, elem_2, operation, table_name);
    break;

case 2: db_head = order_by(db_head, condition, order, table_name);
    break;

case 3: db_head_groupby = group_by(db_head, condition, table_name);
    break;

case 3: db_head_groupby = group_by(db_head, condition, table_name);
    break;
}
```

Infine, se necessario, filtro la colonna e poi stampo il risultato della guery nel file.

```
//rimuovo eventuali colonne non richieste
if (strcmp(columns, "*") != 0 && type!=3)
db_head=filter_columns(db_head, columns, table_name);

//stampo il risultato della query nel file
if (type != 3)
print_query(db_head, str, table_name, columns);
else
print_query_gb(db_head_groupby, str, table_name, condition);

return true;
```

-create:

Questa funzione serve a creare una nuova tabella. Una volta ricevuta la query, ogni parola al suo interno viene ciclata salvando inoltre informazioni come il nome della tabella e le colonne. Infine crea la tabella e ne genera la prima riga.