Algoritmica – Esame di Laboratorio

13/09/2013

Istruzioni

Risolvete il seguente esercizio prestando particolare attenzione alla formattazione dell'input e dell'output. La correzione avverrà in maniera automatica eseguendo dei test e confrontando l'output prodotto dalla vostra soluzione con l'output atteso. Si ricorda che è possibile verificare la correttezza del vostro programma su un sottoinsieme dei input/output utilizzati. I file di input e output per i test sono nominati secondo lo schema: `input0.txt output0.txt input1.txt output1.txt ...` Per effettuare le vostre prove potete utilizzare il comando del terminale per la redirezione dell'input. Ad esempio:

```
./compilato < input0.txt
```

effettua il test del vostro codice sui dati contenuti nel primo file di input, assumendo che `compilato` contenga la compilazione della vostra soluzione e che si trovi nella vostra home directory. Dovete aspettarvi che l'output corrisponda a quanto contenuto nel file `output0.txt`. Per effettuare un controllo automatico sul primo file input `input0.txt` potete eseguire i comandi:

```
./compilato < input0.txt | diff - output0.txt
```

Il comando esegue la vostra soluzione e controlla le differenze fra l'output prodotto e quello corretto.

Una volta consegnata, la vostra soluzione verrà valutata nel server di consegna utilizzando altri file di test non accessibili. Si ricorda di avvisare i docenti una volta che il server ha accettato una soluzione come corretta.

Esercizio

Scrivere un programma che prenda in input:

- un intero N, che rappresenta la dimensione di una tabella hash;
- una stringa S di `unsigned char`, costituita solo da caratteri alfanumerici (a–Z e 0–9, senza spazi) e avente una lunghezza compresa tra i 3 e i 500 caratteri;
- una stringa X di esattamente 3 `unsigned char`.

L'esercizio consiste nell'inserire tutti i 3-grammi distinti della stringa S in una tabella hash per contare le loro frequenze. Un 3-gramma di S è una sequenza di tre carratteri consecutivi che occorre in S. Ad esempio, sia $S =$ `mississippi` e $K = 11$, S contiene i sequenti $K - 2$ 3-grammi non distinti: $S[0,2] =$ `mis`, $S[1,3] =$ `iss`, $S[2,4] =$ `ssi`, $S[3,5] =$ `sis`, $S[4,6] =$ `iss`, $S[5,7] =$ `ssi`, $S[6,8] =$ `sip`, $S[7,9] =$ `ipp`, $S[8,10] =$ `ppi`.

I 3-grammi distinti devono essere inseriti in una tabella hash di dimensione N con conflitti gestiti tramite liste monodirezionali. L'obiettivo è quello di individuare i 3-grammi distinti in S e calcolare la loro frequenza. Il 3-gramma Y dovrà essere inserito in una tabella hash utilizzando la funzione hash $h()$ definita come segue:

$$h(Y) = ((Y[0] \times 256^2 + Y[1] \times 256 + Y[2]) \ \% \ 999149) \ \% \ N$$

dove $Y[0]$ è il primo carattere del 3-gramma, $Y[1]$ il secondo, $Y[2]$ il terzo.

Il programma dovrà scandire S da sinistra verso destra. Per ogni 3-gramma Y incontrato, si dovrà controllare se il 3-gramma Y è già presente nella lista che si trova in posizione $h(Y)$ della tabella. In caso affermativo, si dovrà incrementare la sua frequenza. Se Y non è presente, esso deve essere inserito **in coda** alla lista.
Dopo aver costruito la tabella hash, il programma dovrà stampare tutti i 3-grammi contenuti nella lista in posizione $h(X)$ e le loro frequenze. I 3-grammi dovranno essere stampati, uno per riga, mantenendo il loro ordine nella lista.
NOTA 1: si faccia attenzione a usare sempre il tipo `unsigned char` invece di `char` in modo da avere sempre valori positivi durante il calcolo della funzione hash.
NOTA 2: l'implementazione deve richiedere tempo lineare nella lunghezza di S. Non è quindi ritenuta accettabile una soluzione che calcoli le frequenze dei 3-grammi attraverso scansioni di S.

L'input è formattato nel seguente modo. La prima riga contiene l'intero N, la seconda riga contiene la stringa S, mentre la terza contiene il 3-gramma X.

L'output invece è costituito dai 3-grammi presenti nella lista in posizione $h(X)$ e le loro frequenze, stampati uno per riga.

Esempio

Input

```
5
mississippi
sis
```

Output

```
ssi 2
sis 1
```

Infatti, la tabella hash di dimensione 5 ottenuta per i 3-grammi nella stringa `mississippi`, è

```
0 ipp(1)
1 mis(1) iss(2) ppi(1)
2 ssi(2) sis(1)
3
4 sip(1)
```

e $h(\texttt{ssi}) = 2$.