

Laboratorio di Algoritmi e Strutture Dati

Docenti: M. Goldwurm, S. Aguzzoli

Appello del 18 Settembre 2002

Progetto “Dizionario”

Consegna entro il 9 Ottobre 2002

Il problema

Obiettivo del progetto è la realizzazione di un dizionario e di alcune operazioni sulle parole in esso contenute.

Una *parola* è una sequenza finita di caratteri appartenenti all’alfabeto composto dalle 26 lettere minuscole $\{a, b, c, \dots, z\}$.

Uno *schema* è una sequenza finita di caratteri appartenenti all’alfabeto $\{a, b, c, \dots, z\} \cup \{A, B, C, \dots, Z\}$.

Un *assegnamento* è una funzione σ da $\{A, B, C, \dots, Z\}$ a $\{a, b, c, \dots, z\}$ (ossia, a ogni lettera maiuscola l’assegnamento associa una lettera minuscola).

Dato uno schema $W = \alpha_1 \dots \alpha_n$ e un assegnamento σ , denotiamo con $\sigma(W)$ la parola $w = \beta_1 \dots \beta_n$ tale che, per ogni $1 \leq i \leq n$, se α_i è una lettera maiuscola allora $\beta_i = \sigma(\alpha_i)$, se α_i è una lettera minuscola allora $\beta_i = \alpha_i$.

Una parola w è una *istanza* di uno schema W se esiste un assegnamento σ tale che $w = \sigma(W)$. Ad esempio, “acca” è una istanza dello schema “aBBa”, dello schema “aBCa” e dello schema “CDcC”. Ogni parola di 4 lettere è una istanza dello schema “ABCD”, mentre lo schema “ABBA” identifica le parole palindrome di 4 lettere.

Definiamo *operazioni elementari di editing* su una parola w le seguenti:

- Inserzione di un carattere in qualsiasi posizione in w . Ad esempio, “pippo” diventa “pioppo” tramite inserzione di “o”.
- Cancellazione di un carattere in qualsiasi posizione in w . Ad esempio, “capra” diventa “capa” tramite cancellazione di “r”.
- Sostituzione di un carattere con un altro in qualsiasi posizione. Ad esempio, “cane” diventa “rane” tramite la sostituzione di “c” con “r”.
- Scambio della posizione di due caratteri adiacenti qualsiasi. Ad esempio “trota” diventa “torta” scambiando di posizione “r” con “o”.

Date due parole x e y la loro *distanza di editing* è data dal minor numero di operazioni elementari di editing necessarie per passare da x a y . Ad esempio, “cavolo” e “cavallo” hanno distanza 2, in quanto si passa dalla prima alla seconda sostituendo la prima “o” con una “a” e inserendo una “l”; la distanza fra “capra” e “arpa” è 2, in quanto occorre cancellare la “c” e scambiare la posizione di “p” e “r”; la distanza fra “pesce” e “sedia” è 4 (si noti che ci sono più modi per passare da “pesce” a “sedia” con 4 operazioni elementari).

Si richiede di implementare una struttura dati efficiente che permetta di eseguire le operazioni seguenti. Per quanto riguarda il calcolo della distanza di editing si suggerisce di considerare tecniche di programmazione dinamica.

- **carica (file)**

Inserisce nel dizionario le parole contenute nel file di nome `file`; `file` è di tipo testo e le parole sono separate da uno o più caratteri di spaziatura (compresi tabulatori e newline). Se `file` non esiste non viene eseguita alcuna operazione.

- **visualizza()**

Stampa tutte le parole del dizionario.

- **inserisci (w)**

Inserisce nel dizionario la parola `w`; se `w` è già presente non viene eseguita alcuna operazione.

- **elimina (w)**

Elimina dal dizionario la parola `w`; se `w` non è nel dizionario non viene eseguita alcuna operazione.

- **ricerca (W)**

Stampa `W` e poi la lista di tutte le parole nel dizionario che sono istanze dello schema `W`.

- **distanza (x, y)**

Calcola e stampa la distanza di editing fra le due parole `x` e `y` e inserisce `x` e `y` nel dizionario (se non sono già presenti).

Specifiche di implementazione

Il programma deve leggere dallo standard input (`stdin`) una sequenza di linee (separate da `\n`), ciascuna delle quali corrisponde a una linea della prima colonna della Tabella 1, dove `x, y, w` sono parole e `W` è uno schema e i vari elementi sulla linea sono separati da uno o più spazi. Quando una linea è letta, viene eseguita l'operazione associata; Le operazioni di stampa sono effettuate sullo standard output (`stdout`), e ogni operazione deve iniziare su una nuova linea.

LINEA DI INPUT	OPERAZIONE
<code>c file</code>	carica (file)
<code>v</code>	visualizza()
<code>i w</code>	inserisci (w)
<code>e w</code>	elimina (w)
<code>r W</code>	ricerca (W)
<code>d x y</code>	distanza (x, y)
<code>f</code>	Termina l'esecuzione del programma

Tabella 1: Specifiche del programma

Si noti che non devono essere presenti vincoli sulla lunghezza delle parole e degli schemi.

Esempio

Si supponga che il file `f1.txt` sia:

cane cena cavallo banana ananas dado

e f2.txt sia:

tata alba cute balena cane raro anna
cura baccba

e che le linee di input siano

```
c f1.txt
i cavolo
r cXYZ
c f2.txt
r cXYZ
r ABAC
r baXYZa
r baXYXa
r ABAB
e tata
r ABAB
v
d carota cartone
r caAoBC
f
```

L'output prodotto dal programma deve essere:

```
cXYZ cane cena
cXYZ cane cena cute cura
ABAC dado tata raro
baXYZa banana balena baccba
baXYXa banana
ABAB tata
ABAB
cane cena cavallo banana ananas dado cavolo alba cute
balena raro anna cura baccba
3
caAoBC cavolo carota
```

È irrilevante l'ordine in cui compaiono le parole relative ad ogni operazione.

Presentazione del progetto

Il progetto deve essere inviato per posta elettronica all'indirizzo aguzzoli@dsi.unimi.it entro il 9 Ottobre 2002. La discussione del progetto e l'esame orale si svolgeranno il 14 Ottobre alle 14:30 2002 in luogo da stabilirsi.

Occorre presentare:

1. il codice sorgente (rigorosamente ANSI C, compilabile con **gcc**);

2. una sintetica relazione (formato pdf o rtf) che illustra le strutture dati utilizzate e analizza il costo delle diverse operazioni.

I due file devono essere contenuti in un unico file `.zip` il cui nome dovrà essere `cognome.zip`. La relazione e il codice devono riportare il vostro nome, cognome e matricola.

Una copia cartacea della relazione e del codice deve inoltre essere consegnata al dr. Aguzzoli sempre entro il 9 Ottobre 2002 (lasciandola eventualmente nella sua casella postale presso il dipartimento in via Comelico).

Si ricorda infine di presentarsi alla prova orale con una copia stampata della relazione e del codice.

Per ogni ulteriore chiarimento:

E-mail: `aguzzoli@dsi.unimi.it`

Ricevimento: il mercoledì, ore 15-16, stanza S211.

Avvisi

La versione aggiornata del progetto è pubblicata in `.pdf` sul sito:

`http://homes.dsi.unimi.it/~aguzzoli/stefanodidattica.htm`.

Si consiglia di consultare periodicamente questo sito per eventuali correzioni e/o precisazioni relative al testo del progetto.

Lo svolgimento del progetto è una prova d'esame da svolgere *individualmente*.