# Esercizio

**Scrivere un programma che dati in input un file che contiene un testo e un file che contiene un insieme di parole (una per ogni riga), crei un file in output il cui contenuto è dato dall'insieme delle parole con riportata la frequenza di occorrenza di ogni parola all'interno del testo (in ogni riga una parola seguita dalla relativa frequenza).**

Aiuto 1: partire dalla soluzione all'esercizio visto la volta scorsa.

Aiuto 2: leggere tutto il contenuto del file di testo in una variabile stringa.

Aiuto 3: utilizzare il metodo `split` già visto per ottenere la lista delle parole.

# Prima soluzione (parziale)

```
import sys
punteggiatura = {}       # inizializza dizionario vuoto
punteggiatura['.'] = 0   # inserisce la punteggiatura nel dizionario
punteggiatura[','] = 0
punteggiatura[':'] = 0
punteggiatura[';'] = 0
punteggiatura['\n'] = 0
punteggiatura[')'] = 0
punteggiatura['('] = 0
punteggiatura['\''] = 0
punteggiatura['\"'] = 0
f_in = open(sys.argv[1],"r")  # apre il primo file argomento, contenente il testo da elaborare
testo = f_in.read()           # legge tutto il contenuto del file nella variabile stringa testo
f_in.close()                  # chiude il file
lista_testo = testo.split()   # crea una lista di sottostringhe separate dallo spazio in testo
f_in = open(sys.argv[2],"r")  # apre il secondo file argomento, contenente le parole da cercare
lista_parole = f_in.readlines() # crea una lista di parole seguite dal ritorno carrello
f_in.close()                  # chiude il file
dizionario_parole = {}        # inizializza un dizionario per le parole
for parola in lista_parole :  # per ogni elemento nella lista lista_parole
    dizionario_parole[parola[:len(parola)-1]] = 0 # lo inserisce nel dizionario rimuovendo il \n
for parola in lista_testo :   # per ogni elemento di lista_testo
  if punteggiatura.has_key(parola[-1]) : # se l'ultimo carattere e' di punteggiatura
    if dizionario_parole.has_key(parola[:len(parola)-1]) : # controllo se, senza l'ultimo
                                                           # carattere, appartiene al dizionario
        # se appartiene al dizionario, incremento di uno la frequenza della parola corrispondente
        dizionario_parole[parola[:len(parola)-1]] = dizionario_parole[parola[:len(parola)-1]] + 1
  else : # se l'ultimo carattere non e' di punteggiatura, ripeto quanto sopra considerando tutti
    if dizionario_parole.has_key(parola) : # i caratteri
        dizionario_parole[parola] = dizionario_parole[parola] + 1
f_out = open(sys.argv[3],"w") # apre il terzofile argomento, dove scrivere il risultato
for parola in dizionario_parole.keys() : # per tutte le parole del dizionario
    f_out.write(parola + " " + str(dizionario_parole[parola]) + "\n") # stampa parola e frequenza
f_out.close()  # chiude il file
```

**Purtroppo non funziona sempre: “...ripetitivo),” oppure “...(C-value si ...”**

# Soluzione

**Idea: togliamo prima tutta la punteggiatura! Sostituiamola con uno spazio...**

```
import sys
def rimuovi_punteggiatura(testo):    # rimuove la punteggiatura da testo
    punteggiatura = {}       # inizializza dizionario vuoto
    punteggiatura['.'] = 0   # inserisce la punteggiatura nel dizionario
    punteggiatura[','] = 0
    punteggiatura[':'] = 0
    punteggiatura[';'] = 0
    punteggiatura['\n'] = 0
    punteggiatura[')'] = 0
    punteggiatura['('] = 0
    punteggiatura['\''] = 0
    punteggiatura['\"'] = 0
    testo2 = ''              # inizializza una nuova stringa vuota
    for carattere in testo :
        if not punteggiatura.has_key(carattere) : # se non e' di punteggiatura
            testo2 = testo2 + carattere           # lo aggiunge
        else :
            testo2 = testo2 + " "             # altrimenti aggiunge uno spazio
    return testo2                     # restituisce il testo senza punteggiatura
f_in = open(sys.argv[1],"r")
testo = f_in.read()
f_in.close()
lista_testo = rimuovi_punteggiatura(testo).split() # crea lista parole
f_in = open(sys.argv[2],"r")
lista_parole = f_in.readlines()
f_in.close()
dizionario_parole = {}
for parola in lista_parole :
    dizionario_parole[parola[:len(parola)-1]] = 0
for parola in lista_testo :
    if dizionario_parole.has_key(parola) :
        dizionario_parole[parola] = dizionario_parole[parola] + 1
f_out = open(sys.argv[3],"w")
for parola in dizionario_parole.keys() :
    f_out.write(parola + " " + str(dizionario_parole[parola]) + "\n")
f_out.close()
```

# Altra soluzione

**Il valore del dizionario punteggiatura non è usato: possiamo usare una lista**

```
import sys
def rimuovi_punteggiatura(testo):    # rimuove la punteggiatura da testo
    # inserisce la punteggiatura in una lista
    punteggiatura = ['.',',',':',';','\n',')','(','\'','\"']
    testo2 = ''                      # inizializza una nuova stringa vuota
    for carattere in testo :
        if not carattere in punteggiatura : # se non e' di punteggiatura
            testo2 = testo2 + carattere     # lo aggiunge
        else :
            testo2 = testo2 + " "       # altrimenti aggiunge uno spazio
    return testo2                  # restituisce il testo senza punteggiatura
f_in = open(sys.argv[1],"r")
testo = f_in.read()
f_in.close()
lista_testo = rimuovi_punteggiatura(testo).split() # crea lista parole
f_in = open(sys.argv[2],"r")
lista_parole = f_in.readlines()
f_in.close()
dizionario_parole = {}
for parola in lista_parole :
    dizionario_parole[parola[:len(parola)-1]] = 0
for parola in lista_testo :
    if dizionario_parole.has_key(parola) :
        dizionario_parole[parola] = dizionario_parole[parola] + 1
f_out = open(sys.argv[3],"w")
for parola in dizionario_parole.keys() :
    f_out.write(parola + " " + str(dizionario_parole[parola]) + "\n")
f_out.close()
```