

frequenza-di-parole

June 5, 2023

0.1 Occorrenze di parole

In questo esercizio vogliamo calcolare il numero di occorrenze delle parole contenute in un testo. Per esempio dato il testo

Quanta legna taglia il taglia-legna, se ha voglia di tagliare legna?
Legna da ardere, si intende.

Si vorrebbe come risultato una cosa del genere

```
{ 'quanta': 1, 'legna': 4, 'taglia': 2, 'il': 1, 'se': 1, 'ha': 1,  
  'voglia': 1, 'di': 1, 'tagliare': 1, 'da': 1, 'ardere': 1,  
  'si': 1, 'intende': 1}
```

Per fare questo esercizio è utile utilizzare il metodo `split` delle stringhe. Data una stringa possiamo spezzettarla in parole con questo metodo, che restituisce la liste delle parole ottenute.

```
[4]: x="Questa è, senza dubbio, una grande giornata!"  
     x.split()
```

```
[4]: ['Questa', 'è,', 'senza', 'dubbio,', 'una', 'grande', 'giornata!']
```

```
[5]: y=" Casa casa, casa"  
     y.split()
```

```
[5]: ['Casa', 'casa,', 'casa']
```

Una volta divise le parole in una lista possiamo produrre il dizionario con i risultati.

0.1.1 Primo tentativo

```
[7]: def occorrenze(testo):  
     parole = testo.split()  
  
     risultati = {}  
     for parola in parole:  
         if parola in risultati:  
             risultati[parola] = risultati[parola] + 1  
         else:  
             risultati[parola] = 1  
     return risultati
```

```
[8]: x = "Questa è, senza dubbio, una grande giornata!"  
print(occorrenze(x))
```

```
{'Questa': 1, 'è,': 1, 'senza': 1, 'dubbio,': 1, 'una': 1, 'grande': 1,  
'giornata!': 1}
```

```
[9]: y = "    Casa casa,    casa"  
print(occorrenze(y))
```

```
{'Casa': 1, 'casa,': 1, 'casa': 1}
```

```
[11]: z = "Quanta legna taglia il taglia-legna, se ha voglia di tagliare legna? ̀  
↳Legna da ardere, si intende."  
print(occorrenze(z))
```

```
{'Quanta': 1, 'legna': 1, 'taglia': 1, 'il': 1, 'taglia-legna,': 1, 'se': 1,  
'ha': 1, 'voglia': 1, 'di': 1, 'tagliare': 1, 'legna?': 1, 'Legna': 1, 'da': 1,  
'ardere,': 1, 'si': 1, 'intende.': 1}
```

Quest'averione preliminare mette in luce **due problemi**:

1. la funzione `split` separa lungo gli spazi, e quindi la punteggiatura rimane attaccata ad alcune parole. Quindi `casa` e `casa,` vengono considerate parole diverse.
2. due occorrenze della stessa parola, con maiuscole diverse, vengono interpretate come parole diverse.

Esercizio: il secondo problema è il più semplice da risolvere. Si modifichi il programma precedente per farlo. (*Indizio:* usate il metodo `lower`).

0.1.2 Pulizia del testo

Per risolvere il primo problema è possibile fare una pulizia di tutti i caratteri nel testo che non fanno parte di parole (ed esempio la punteggiatura). Un modo abbastanza semplice e corretto potrebbe essere - sostituire tutti i caratteri non alfabetici con degli spazi, - far fare a `split` il resto del lavoro.

```
[13]: def pulizia(testo):  
    res = ""  
    for c in testo:  
        if c.isalpha():  
            res = res + c  
        else:  
            res = res + ' '  
    return res
```

```
[14]: x = "Questa è, senza dubbio, una grande giornata!"  
print(pulizia(x))
```

```
Questa è  senza  dubbio  una  grande  giornata
```

```
[15]: y = "    Casa casa,    casa"
      print(pulizia(y))
```

Casa casa casa

```
[16]: z = "Quanta legna taglia il taglia-legna, se ha voglia di tagliare legna? ┘
      ↪Legna da ardere, si intende."
      print(pulizia(z))
```

Quanta legna taglia il taglia legna se ha voglia di tagliare legna Legna da
ardere si intende

0.1.3 Mettete tutto insieme

Adesso mettete insieme tutto quello che avete imparato e risolvete l'esercizio

```
[ ]: def occorrenze(testo):
      return
```

```
[ ]: x = "Questa è, senza dubbio, una grande giornata!"
      print(occorrenze(x))
```

```
[ ]: y = "    Casa casa,    casa"
      print(occorrenze(y))
```

```
[ ]: z = "Quanta legna taglia il taglia-legna, se ha voglia di tagliare legna? ┘
      ↪Legna da ardere, si intende."
      print(occorrenze(z))
```