

tidytext

Pacchetto per il text mining con R, usando i pacchetti del *tidyverse*, e dunque [dataframe tibble](#), anziché corpora (per molte analisi successive, dovranno comunque essere costruite le [matrici testuali](#)).

Vantaggi:

- l'organizzazione dei dati in questo formato rende molto semplici le operazioni di gestione e trasformazione della base dati, specialmente con i pacchetti del *tidyverse*;
- risulta essere l'opzione più semplice per il text mining, anche grazie ai [pacchetti Tidymodels](#) ;
- compatibile e progressivamente integrato con [Quanteda](#)

Svantaggi:

- non adatto all'analisi testuale e linguistica.

Vedi:

- [aiuto sul pacchetto](#),
- [Text Mining with R!](#) (J. Silge e D. Robinson)
- 📖 [Strumenti per l'analisi testuale e il text mining con R](#) (A. Vardanega, in progress, testo liberamente accessibile da web)

workflow e funzioni

All'interno del testo citato sopra, ho - per il momento - dedicato spazio ai seguenti argomenti:

- 📖 [workflow](#): sintesi dei principali step operativi.
- 📖 [importazione dei testi](#); vedi anche, qui, la voce [Importazione dei testi](#)
- 📖 [segmentazione](#) dei documenti
- 📖 [tokenizzazione](#)
- 📖 [ngrams](#)
- 📖 [statistiche riassuntive del corpus](#) con *dplyr*
- 📖 [frequenze](#) con *dplyr*
- 📖 [conversione di corpora in dataframe con la funzione tidy](#)

[Analisi testuale](#), [pacchetti](#), [text mining](#), [tidyverse](#)

From:

<https://www.agnesevardanega.eu/wiki/> - **Ricerca Sociale con R**

Permanent link:

<https://www.agnesevardanega.eu/wiki/r/analisi-testuale/tidytext>

Last update: **13/08/2025 12:11**

