

Guida VOYANT TOOLS



a cura di Sonia Grassia

(PRIMO RIQUADRO A SINISTRA)

1. CIRRUS: word cloud dove abbiamo la rappresentazione grafica del “vocabolario” delle parole che ricorrono nel testo; posizionando il mouse su ogni parola possiamo anche vedere la Frequenza assoluta (quante volte appare effettivamente la parola nel testo).
2. TERMINI: ci dà la lista delle parole, il “vocabolario” del nostro testo; possiamo ordinarle in ordine alfabetico o sulla base della Frequenza assoluta (Conteggio). Cliccando sulla freccia accanto alla colonna “Parola” abbiamo la possibilità di aggiungere ulteriori colonne, quella che può interessare a noi è “Relativa” in cui verranno mostrate le Frequenze relative (quindi normalizzate per milione).
3. LINK: è il Grafico delle Collocazioni in cui le parole più frequenti sono collegate ai rispettivi collocati (arancioni) (sono quelle parole che occorrono frequentemente con una parola-chiave (blu)); se puntiamo con il mouse ai collocati si aprirà una nuvola che ci mostra la frequenza nel contesto di prossimità; con il tasto sinistro del mouse abbiamo la possibilità di rimuovere, rendere parola-chiave, aggiungere altri collocati.

(SECONDO RIQUADRO-CENTRALE)

4. LETTORE: ci mostra il testo che abbiamo caricato; la barra blu in basso a questo riquadro ci fornisce, dopo aver cliccato su una parola, informazioni che riguardano la frequenza e la distribuzione di essa all'interno del testo.
5. TERMSBERRY: è un grafico che, oltre alla frequenza delle parole, considera anche i collocati (diciamo che è un ibrido tra Cirrus e Link).

(TERZO RIQUADRO-DESTRA)

6. ANDAMENTI: è un grafico lineare delle Frequenze relative (la divisione del documento in sezioni avviene in automatico); dopo aver caricato il

documento compariranno in automatico le parole con una Frequenza assoluta maggiore, successivamente possiamo cercare (in basso alla finestra a sinistra) le parole che vogliamo vedere nel grafico, oppure selezionarle dal riquadro “Termini”; sempre nel medesimo riquadro, in basso a destra, abbiamo “Display” dove possiamo scegliere diverse tipologie di grafici.

7. PAROLE DEL DOCUMENTO: è una tabella delle parole del documento che possono essere ordinate (anche qui) alfabeticamente, secondo la Frequenza assoluta (Conteggio) o quella relativa; inoltre abbiamo un grafico Sparkline, inerente ad ogni singola parola, che ci mostra l’andamento della frequenza del termine nel testo; cliccando sulla freccia che appare se posizioniamo il puntatore sopra un’altra colonna, abbiamo la possibilità di aggiungere altre colonne, quella che potrebbe essere interessante per noi è “Z-Score” (valore che mostra quanto è frequente e distribuito un termine rispetto ad altri all’interno del testo).

(PRIMO RIQUADRO-in basso-A SINISTRA)

8. SOMMARIO: il “totale parole” corrisponde ai tokens, sono tutte le parole conteggiate nel testo, invece le “forme di parola uniche” sono le parole-tipo (i type) del testo, cioè tutte le parole del testo ma conteggiate una sola volta nel caso in cui dovessero ripetersi; “Densità del vocabolario” ci dice la ricchezza lessicale del testo, sostanzialmente è la divisione tra “forme di parola uniche” e “totale parole”; “Average Words Per Sentence” ci mostra la paratassi o ipotassi (quante parole ci sono mediamente in una frase).
9. DOCUMENTI: la sezione documenti ti permette di aggiungere un secondo testo, riportando poi nella sezione SOMMARIO i dati della comparazione. Per aggiungere il secondo documenti si deve cliccare MODIFICA, in basso, e poi AGGIUNGI. Da qui possiamo incollare il testo e cliccare INVIO.

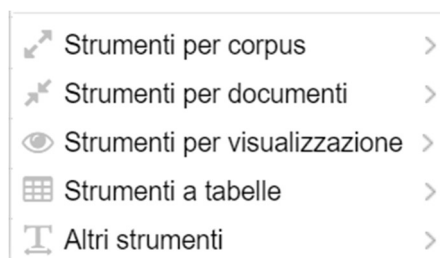
(SECONDO RIQUADRO-in basso-A DESTRA)

10. CONTESTI: ci mostra le occorrenze di ciascuna parola-chiave (Parola) e il contesto in cui essa appare (Sinistra e Destra); può essere utile per comprendere come alcune parole vengono utilizzate in determinati contesti, anche diversi tra loro, o se ad esempio sono ricorrenti in contesti simili.
11. BUBBLELINES: è un altro grafico nel quale ogni parola è rappresentata sottoforma di bolla, quanto più grande è la dimensione di quest’ultima tanto sarà maggiore la Frequenza del termine stesso.

12. CORRELATION: ci dice la misura in cui le Frequenze dei termini variano in sincronia(ci dice se le frequenze dei due termini sono direttamente o inversamente proporzionali); se il valore di “Correlation” è uguale a 1 significa che le frequenze dei due termini sono direttamente proporzionali (se aumenta la frequenza di un termine, aumenterà anche quella del secondo), al contrario, se è uguale a -1 significa che le frequenze dei due termini sono inversamente proporzionali(all’aumentare della frequenza del termine 1 diminuirà quella del termine 2 e viceversa); Il valore “Significance” sta ad indicare la significatività di una correlazione tra due termini: se questo valore è uguale o inferiore allo 0.05 significa che i due termini si trovano in stretta correlazione (ATTENZIONE: usate con cautela questo valore perché dipende dal valore della distribuzione normalizzata e visto che noi utilizzeremo un testo breve con valori di frequenza bassi può sembrare un valore statisticamente rilevante ma solo perché abbiamo a che fare con una piccola popolazione → NON VE LO SO SPIEGARE BENE, SORRY)

ALTRI TOOLS

In ciascun riquadro abbiamo la possibilità di poter cambiare il tools: accanto al “?” vi è una finestra simile al logo di windows, cliccando apparirà una finestra:



Alcuni potranno essere utili per la nostra analisi, tipo:

1. COLLOCAZIONI: (si trova in “strumenti per corpus”) vi sono i termini che compaiono più frequentemente in prossimità delle parole chiave nel corpus (“Parole” sono le parole-chiavi, “Collocazione” è il collocato che compare vicino alla parola-chiave, “Conteggio(contesto)” è il numero di occorrenze di questa collocazione accanto alla parola-chiave).

Altri sono rappresentazioni differenti dello stesso fenomeno, quindi decidete voi quale utilizzare (tanto non sanno nemmeno dove stanno di casa e dovremo comunque spiegarglielo noi a cosa servono tutte queste minchiate che copieremo e incolleremo).

!!! Infine non dimenticate che, qualsiasi tipo di valutazione statistica in testi brevi non può darci la possibilità di sfruttare al meglio tutti questi strumenti perché non ci forniscono dati statisticamente rilevanti, al massimo possono confermare, attraverso dei dati statistici, ciò che si può dedurre a seguito di una lettura e comprensione di un testo.

Ciao Patasfronzoli, vi voglio bene!<3

Sonia 😊