

## Normalizzare DB con l'aiuto dell'algoritmo Soundex

**Author :** riccardo

**Categories :** [Domande Lettori](#), [Macro](#)

**Date :** 7 agosto 2013

Fabio ci scrive

“storicamente nell’arco di alcuni decenni, differenti persone in differenti modi hanno raccolto dati. La raccolta ammonta ad un totale di circa 3.000 celle all’anno per ogni foglio ed è un incubo perchè lo stile dei fogli è stato stravolto dall’avvicinarsi eterogeneo di decine di collaboratori diversi che si sono succeduti nello stesso ufficio a raccogliere i dati. Dati inseriti in maiuscolo, minuscolo, scritti correttamente per intero, abbreviati, storpiati e/o alterati... cut....dovrei scoprire il modo di far cercare ad Excel variabili più o meno farcite di wildcard singole tipo le '?' o '\*' da vecchio DOS. Per dire: Corso Vittorio Emanuele II lo posso per esempio trovare come CORSO, corso, C.so, C; Vittorio, VITTORIO, V.; Emanuele, Emmanuele o Emanuele ... una cosa da uscire pazzi. Un Vs. collega dal suo blog mi suggeriva: lo metterei tutti i dati in un unico foglio. Ognuno in una sua riga. La prima colonna conterrà l’anno, così non si perde il riferimento.

Poi, la ricerca ne viene probabilmente facilitata anche con l’uso dell’algoritmo Soundex.

.... cut....l’ho incollata in un modulo ma non riesco a trovare il modo di applicarla.”

Risposta:

Personalmente non conoscevo questo algoritmo e ne ho trovato una definizione al seguente link:

[http://www.excelvba.it/Forum/story/Visual\\_Basic\\_for\\_Applications/Soundex.html](http://www.excelvba.it/Forum/story/Visual_Basic_for_Applications/Soundex.html)

Soundex è un algoritmo di fonetica per l’indicizzazione di parole secondo il loro suono o pronuncia. cut... L’obiettivo è di poter raggruppare parole omofone (dal suono simile) affinché possano essere abbinate per un ricerca veloce, nonostante differenze o errori di ortografia. cut...Ad esempio, MATTONE è codificato come M350; MATTINO e METANO hanno la stessa codifica M350 (e infatti hanno un suono simile)

L’algoritmo non è in grado di associare le stringhe contenenti abbreviazioni come “Corso” e “C.so”

Una volta inserito l’algoritmo all’interno di un Modulo, posto di avere una parola nella cella A1,

nella cella B1 basta scrivere =soundex(A1,8) dove 8, a scelta dell'utilizzatore, indica il grado di precisione desiderato (la lunghezza del codice restituito).

L'algoritmo non salva dalla lunga opera manuale di normalizzazione dei dati, ma da certamente un aiuto. Ecco come ho proceduto per automatizzare un pò l'operazione.

**ATTENZIONE:**

per vostra sicurezza create una copia del file da analizzare in modo da essere sicuri di avere la base dati originale.

>>> l'azione delle macro non è reversibile