

Selezioni per Somiglianza di Termini

Si considerano somiglianti termini che differiscano gli uni dagli altri per un certo numero di caratteri in più, in meno o semplicemente diversi. Il grado di somiglianza va quindi regolato per ottenere il risultato ottimale.

Come ragiona il Server

eXtraWay Server si basa su 3 fattori

Fattore	Significato
Prefisso	Rappresenta il numero di caratteri all'inizio del termine che <u>deve</u> risultare uguale. Esso può essere nullo o un numero ragionevolmente piccolo ¹⁾ . L'uso di un prefisso causa una riduzione della capacità del server di identificare chiavi simili a quelle date ma diviene indispensabile specialmente quando la somiglianza viene richiesta su campi testuali appartenenti ad archivi di ampie dimensioni. Il tentativo di estendere per somiglianza una chiave su un simile vocabolario potrebbe comportare tempi di elaborazione eccessivamente dilatati senza riuscire quindi a dare una risposta in tempo utile. Per tale ragione, quando un archivio ha una simili dimensioni può essere configurato per forzare un prefisso di default . Solo le chiavi che sono del tutto uguali nel prefisso entreranno a far parte del test di dettagli sulla somiglianza.
Caratteri & Errori	Questi due fattori vanno visti insieme. In sostanza il grado di somiglianza è dato, comprendendo anche il <i>Prefisso</i> , sulla base di quanti <i>Errori</i> sono ammessi su quanti <i>Caratteri</i> e multipli di caratteri.

Cerchiamo di dare un senso più chiaro alle indicazioni date. Le regole di somiglianza, intese come *Errori* su un certo numero di *Caratteri* si applicano comprendendo il *Prefisso* che entra a far parte del calcolo pur imponendo la sua uguaglianza esatta. Facciamo un esempio "classico": ammettiamo 1 errore su 4 caratteri.

Questo comporta un ragionamento che si può riassumere come segue e che non dipende dal *Prefisso*.

Caratteri	Comportamento
Da 1 a 3	Viste le dimensioni, non viene ammesso alcun errore. Le chiavi più brevi della soglia minima devono essere tutte pienamente corrispondenti per essere considerate valide. Fare una richiesta per una somiglianza indicando un termine più breve della soglia di caratteri richiesta rende, di fatto, vana l'estensione alla somiglianza perché solo il termine esatto verrà eletto.
Da 4 a 7	La chiave, una volta valutato l'eventuale <i>Prefisso</i> , verrà considerata somigliante se presenta 1 solo errore ²⁾ .
Da 8 a 11	Date le dimensioni, la chiave si considererà somigliante se presenta 2 errori.

In altre parole, assumendo *c* il numero dei caratteri sui quali sono consentiti e errori, gli errori totali ammessi per una singola chiave del vocabolario saranno pari alla sua lunghezza $\frac{c}{e}$.

Espressione del Comando

Compito del chiamante è esplicitare i fattori indicati, ovvero il *Prefisso*, il numero di *Caratteri* ed il numero di *Errori* per i caratteri indicati.

Se il comando viene effettuato utilizzando le interfacce canoniche³⁾ in esso possono essere indicati tutti e 3 i fattori.

Alternativamente è possibile⁴⁾ indicare direttamente i fattori in fase di selezione seguendo la sintassi che segue.

[nomecampo| (Fuzzy:<prefix>.<chars>)] = termine

L'esempio mette in evidenza la necessità di aggiungere un modificatore di ricerca denominato Fuzzy che prevede un parametro rappresentato da un numero decimale. In esso, la parte intera rappresenta il *Prefisso*, la parte decimale il numero di *Caratteri* mentre il numero di *Errori* si considera fissato a 1⁵⁾.

Valori di Default

Il server prevede un valore di default per il prefisso. Come detto in precedenza, se un vocabolario è particolarmente vasto può risultare indispensabile imporre un restringimento e richiedere che almeno una parte del prefisso del termine sia del tutto corrispondente.

L'applicazione che interroga il server ha quindi facoltà di indicare un prefisso e se esso viene indicato il server rispetterà quanto richiesto.

Se per contro il prefisso non viene indicato⁶⁾, il server verificherà la presenza di un eventuale prefisso di default e lo applicherà. Si veda la [documentazione relativa](#) alla voce [search.similarprefix](#).

¹⁾

Di solito tra 1 e 3

²⁾

Si ricorda che l'errore si intende un carattere in più, in meno o differente.

³⁾

Vds. documentazione del Broker

⁴⁾

A partire dalla versione 23.0.0.0 del server eXtraWay

⁵⁾

La regolazione della somiglianza può facilmente essere ottenuta modificando semplicemente il numero di *Caratteri* ottenendo risultati corrispondenti alla regolazione di entrambe i fattori.

⁶⁾



Vale a dire se viene indicato pari a '0'