MI

Massimo Ivaldi

A SEARCH COMPANY

WHITE PAPER / RICERCA SEMANTICA · E-COMMERCE

Ricerca semantica per e-commerce, un caso studio

Da 36 query senza risposta a zero

Dati, metodo e limiti di un intervento reale su un catalogo tecnico.



settembre 2026 · white paper

LA RICERCA COME SISTEMA DI CRESCITA.

Dal lessico al significato.

Un caso reale su 1.535 prodotti mostra come dati migliori e ricerca ibrida possano eliminare le query senza risposta e ridurre drasticamente i tempi peggiori.

1.535	121	36 → 0	3.237 → 168
PRODOTTI NEL CATALOGO	QUERY DI PROVA	QUERY SENZA RISPOSTA	MILLISECONDI NEL CASO PEGGIORE

DENTRO QUESTO CASO STUDIO

01 **Abstract**

02 **Che cos'e' la ricerca semantica, ossia per significato**

03 **Il metodo: prima i dati, poi la misura**

04 **L'arricchimento del catalogo**

05 **Il corpo di prova: come si misura una ricerca**

06 **La misura di partenza: la ricerca nativa di PrestaShop**

07 **Il passaggio a Meilisearch**

08 **I sinonimi: una verifica che cambia l'approccio**

09 **Il risultato**

10 **Che cosa questi numeri non dicono**

11 **In sintesi**

Il risultato centrale

A parità di catalogo e di 121 query, le ricerche senza risposta passano da 36 a zero. Il tempo nel caso peggiore scende da 3.237 a 168 millisecondi.

ABSTRACT

Abstract

La ricerca interna di un negozio online e' spesso il punto in cui si perdono i clienti che avevano gia' deciso di comprare. Questo documento descrive un intervento completo su un catalogo di utensileria di millecinquecentotrentacinque articoli, condotto in due tempi: prima la ricostruzione dei dati di prodotto, poi la sostituzione del motore di ricerca con uno capace di cercare per significato e non solo per corrispondenza di lettere.

Il lavoro e' stato misurato con un insieme di centoventuno domande di prova, in parte costruite per gradi di difficolta' crescente, in parte raccolte da quello che le persone digitano davvero sui motori di ricerca. Le stesse domande sono state poste al motore vecchio e a quello nuovo, a parita' di catalogo.

Il risultato: le domande che non ricevevano nessuna risposta passano da trentasei a zero, e il tempo di attesa nel caso peggiore scende da tre secondi e due a centosessantotto millesimi. Il documento riporta anche quello che la misura non dimostra, e i punti in cui la ricerca semantica, usata senza accortezze, produce errori peggiori di quelli che risolve.

L'intervento e' stato realizzato su PrestaShop perche' era la piattaforma del cliente. Il metodo non dipende dalla piattaforma: gli stessi passaggi si applicano a WooCommerce, Magento, Shopify e a un e-commerce sviluppato su misura. Cambiano i nomi dei campi e il modo di aggiungerli, non la sostanza del lavoro.

Un nostro cliente vende utensileria e materiale da lavoro con un negozio PrestaShop. Millecinquecentotrentacinque articoli: trapani, avvitatori, batterie, pinze, viteria, smerigliatrici. Un catalogo piccolo ma tecnico, dove ogni pezzo ha misure, tensioni, materiali e compatibilita'.

Il problema, come ci e' stato esposto, e' quello che sentiamo formulare quasi sempre con le stesse parole: "la ricerca del sito non funziona". Chi cerca non trova, e chi non trova se ne va.

Dietro quella frase ci sono sempre tre cause che si sommano: un motore di ricerca costruito per ritrovare cio' che si sa gia' come si chiama, un catalogo scritto nella lingua del produttore invece che in quella del cliente, e domande che descrivono un problema invece di nominare un prodotto. Le tre cause si aggravano a vicenda, e nessuna delle tre si risolve da sola.

Questo documento racconta che cosa abbiamo trovato, che cosa abbiamo fatto e quanto si e' spostato il risultato, con i numeri.

CASO STUDIO

Che cos'e' la ricerca semantica, ossia per significato

Una ricerca tradizionale confronta lettere. Se scrivi "flessibile" e nel catalogo quella parola non c'e', non trova niente, anche se il negozio e' pieno di smerigliatrici angolari, che in ferramenta si chiamano proprio cosi'.

La ricerca per significato funziona in un altro modo. Ogni scheda prodotto viene trasformata in una lunga lista di numeri, che si chiama **vettore**, e che rappresenta il senso del testo. La domanda del cliente viene trasformata allo stesso modo. Poi il motore cerca le schede il cui vettore assomiglia di piu' a quello della domanda. Il calcolo che trasforma il testo in numeri si chiama **embedding**, e lo fa un modello di intelligenza artificiale addestrato apposta.

E' questo che permette a "attrezzo per spelare i fili" di arrivare a una pinza spellacavi senza condividere una sola parola con la scheda.

Che cosa risolve

Il vantaggio non e' che trova di piu': e' che risponde a chi **non sa come si chiama la cosa che vuole**. Chi conosce il codice dell'articolo lo trovava anche prima, e chi conosce il nome tecnico pure. La ricerca semantica serve a tutti gli altri, che sono la maggioranza, e che oggi se ne vanno a mani vuote.

Tre esempi presi dalle misure di questo lavoro, tutti con la ricerca esistente a zero risultati.

Chi descrive il problema invece di nominare l'oggetto.

> `cosa serve per forare il muro  ->  Martello perforatore AEG BH 24E SDS-plus`

Nella scheda di quel martello non compaiono ne' "cosa", ne' "serve", ne' "forare", ne' "muro". Compaiono "foratura", "muratura", "calcestruzzo". Il modello sa che sono la stessa area di senso.

Chi parla come parlerebbe a un commesso.

> `vorrei un avvitatore leggero da usare in casa  ->  Avvitatore elettrico AEG S 4000 E`

Chi usa il nome del mestiere invece di quello del produttore.

> `attrezzo per spelare i fili  ->  Pinza spellacavi Microconnect PP-LSA2`

La parola "spelare" nel catalogo non esiste: esistono "spellafili" e "rimuovere la guaina".

Su questo catalogo le domande di questo tipo erano circa un terzo del totale, e la ricerca esistente non ne serviva nessuna.

Dove sbaglia, e perche' va governata

Il vettore non sa di non sapere.

Una ricerca a parole, quando non trova, restituisce zero risultati: e' spiacevole ma onesto, e il cliente capisce che quella cosa li' non c'e'. Una ricerca per significato restituisce **sempre** qualcosa, perche' fra tutti i prodotti ce n'e' sempre uno piu' vicino degli altri. A volte e' vicino per il motivo sbagliato.

Tre casi, tutti presi dalle prove con la ricerca puramente semantica.

Segue il suono della parola invece del senso.

> `tronchese  ->  Troncatrice AEG PS 216 L, 1.500 W`

Il cliente cercava una pinza tagliacavi da tasca, gli viene proposta una macchina da banco. Due oggetti che non hanno in comune nulla tranne le prime cinque lettere.

Sbaglia il codice articolo, e lo sbaglia di poco.

```
WZ0060   ->   Pinza LogiLink WZ0068
```

Questo e' il piu' insidioso, perche' il risultato **sembra giusto**: il cliente ordina, riceve il pezzo sbagliato, e la colpa se la prende il negozio. Nelle prove la ricerca puramente semantica ha sbagliato quattro codici su cinque.

Associa in modo sensato e sbagliato.

Su un altro catalogo, cercando "salvavita", ha proposto quattro gruppi di continuita'. Il ragionamento si capisce, un apparecchio che salva la corrente, ma il cliente cercava l'interruttore differenziale del quadro elettrico.

Da qui una regola che vale per chiunque valuti un motore di ricerca: **il conteggio di quante ricerche danno almeno un risultato non e' mai una misura sufficiente.** Un motore che sbaglia con sicurezza vince quel conteggio contro un motore che ammette di non sapere.

E da qui la conseguenza pratica, che e' la parte del mestiere: la ricerca semantica non si mette in produzione da sola. Si mescola con quella a parole, in una proporzione che va scelta e misurata sul catalogo di quel negozio.

Due impianti semantici non danno lo stesso risultato

"Ricerca semantica" e' un'etichetta che copre esiti molto diversi fra loro. Due cose in particolare vengono date per scontate e non lo sono.

Gli embedding non sono tutti uguali. I modelli che trasformano il testo in numeri sono molti, di produttori diversi, e differiscono su tre assi che contano tutti.

la lingua	quasi tutti sono addestrati soprattutto sull'inglese. Un modello che capisce bene "angle grinder" puo' capire molto peggio "smerigliatrice angolare", e quasi per niente "flessibile"
il settore	il linguaggio della ferramenta, con capicorda, faston, puntalini e teste svasate, e' lontano dall'italiano su cui i modelli si allenano
costo e dimensione	un modello piu' grande produce vettori piu' lunghi, costa di piu' a generare e occupa piu' spazio nell'indice, e non sempre risponde meglio

Su un catalogo italiano e tecnico, provare due o tre modelli e confrontarli sulle stesse domande e' la differenza fra un impianto che funziona e uno che sembra funzionare.

Il modello va alimentato nel modo giusto. Un motore semantico rende quanto vale il modello che monta e quanto e' curato il testo che gli si consegna: sono due parti dello stesso impianto, e vanno costruite insieme. E' la parte invisibile del lavoro, perche' chi guarda un sito non ha modo di sapere che cosa sia stato dato in pasto al modello quando l'indice e' stato costruito.

CHE COSA SI MANDA	CHE COSA SUCCEDE
solo il titolo	il vettore rappresenta poco, e su titoli fatti di sigle non rappresenta nulla
tutto il documento, codici e prezzi compresi	il vettore si annacqua: un codice articolo non ha un significato da rappresentare, e' rumore
una selezione ragionata	il vettore rappresenta davvero il prodotto

CASO STUDIO

Il metodo: prima i dati, poi la misura

La tentazione, arrivando, e' misurare subito. E' la cosa sbagliata, e per una ragione tecnica prima che metodologica.

Misurare un motore di ricerca su un catalogo dove i titoli non dicono che cosa sia il prodotto non misura il motore: misura il catalogo. Un motore, per quanto buono, non puo' trovare una parola che nel catalogo non e' scritta da nessuna parte. Partendo dai dati sbagliati si otterrebbero numeri bassi per tutti e due i motori, e si finirebbe per attribuire a uno il difetto dell'altro.

Quindi prima abbiamo ricostruito i dati, poi abbiamo misurato. E abbiamo misurato i due motori **a parita' di catalogo**, dando a ciascuno le stesse informazioni, ognuna scritta nel campo che quel motore sa leggere: e' l'unico modo perche' il confronto regga a chi lo contesta.

L'ordine dei due tempi non dice nulla sul loro peso. La ricostruzione dei dati e' il presupposto che permette al motore nuovo di rendere quello che puo' rendere; ma da sola, come si vedra' nei numeri, recupera due domande su centoventuno, mentre il cambio di motore le recupera tutte e trentasei.

CASO STUDIO

L'arricchimento del catalogo

Che cosa non andava

DIFETTO	QUANTI PRODOTTI
titolo che non nomina l'oggetto	192 su 1.535, il 12,5%
descrizione che ripete i dati senza aggiungere nulla	circa il 90%
descrizione scritta da un umano	157, il 10%

Un esempio di titolo: **"AEG BBH 12"**. E' un martello perforatore, ma il titolo non lo dice, e la parola "martello" nella scheda non compare da nessuna parte. Chi cerca "martello perforatore AEG" ci arriva per caso.

Le descrizioni erano elenchi puntati generati dai dati: "Dispositivo a corrente residua, Numero di poli: 1, IP20, Bianco". Non aggiungono una parola a quelle che stanno gia' nelle caratteristiche tecniche.

Come lo abbiamo sistemato

Abbiamo dato a un modello di intelligenza artificiale, prodotto per prodotto, **il record completo della scheda e la fotografia**, chiedendogli tre cose: un titolo nella forma nome dell'oggetto, marca, modello; una descrizione lunga discorsiva che contenga tutte le caratteristiche tecniche ma dica anche a che serve l'oggetto, chi lo usa e in che lavori; una descrizione breve.

La fotografia conta piu' di quanto si creda. Su uno spellafili il modello ha scritto "presenta una struttura a pinza con ganasce regolabili e una vite di registro laterale": nel record quel dettaglio non c'era, l'ha visto nell'immagine.

Il punto critico: il modello inventa

Questa e' la parte che distingue un lavoro fatto bene da un disastro.

Il testo delle istruzioni diceva chiaramente di non inventare nessun dato. **Il modello ha inventato lo stesso**: ha scritto quattordici codici a barre inesistenti, tredici cifre, formato corretto, su prodotti che il codice a barre non ce l'hanno.

Non si risolve riscrivendo le istruzioni. Si risolve con controlli automatici a valle, prima di scrivere qualunque cosa nel catalogo. I nostri erano sei: struttura della risposta valida, titolo entro la lunghezza massima e contenente marca e modello, descrizioni entro i limiti, punteggiatura normalizzata, e **nessun numero di tredici cifre che non sia il codice a barre vero del prodotto**.

Quello che non passa viene riprovato, e se non passa neanche allora viene messo da parte e non scritto.

CASO STUDIO

Il risultato

schede riscritte	1.535 su 1.535
giri necessari	4
codici a barre inventati e bloccati	14
costo totale del modello	circa 1 dollaro
vocabolario della ricerca interna	da 9.001 a 16.055 parole

Il vocabolario quasi raddoppiato e' il dato piu' eloquente: sono le parole che prima non esistevano in nessun punto del catalogo, e che adesso rendono trovabili i prodotti.

CASO STUDIO

Il corpo di prova: come si misura una ricerca

Per dire che una ricerca e' migliorata servono delle domande, e quelle domande non possono essere inventate da chi fa la misura, altrimenti si finisce per scrivere query nella lingua del catalogo, che funzionano sempre.

Abbiamo costruito il corpo di prova da due fonti.

Cinquantacinque query costruite su dieci gradini di difficolta' crescente, dal nome secco alla frase parlata:

GRADINO	ESEMPIO
1. il nome della cosa	`trapano`
2. nome e marca	`martello perforatore AEG`
3. nome e attributo	`trapano a percussione`
4. attributo numerico	`batteria 18v`
5. codice	`BBH18-LI402C`
6. nome popolare	`flessibile , brugola`
7. scopo invece che oggetto	`attrezzo per spelare i fili`
8. categoria larga	`elettroutensili`
9. piu' vincoli insieme	`trapano a batteria 18v leggero`
10. frase parlata	`mi serve un trapano per fare buchi nel cemento senza filo`

Circa cinquanta frasi vere, raccolte dall'autocompletamento di Google. Il procedimento: si prendono i nomi delle categorie e delle marche del catalogo, si danno in pasto all'autocompletamento, e tornano le frasi che le persone digitano davvero. Da 253 semi sono usciti 1.407 suggerimenti.

In totale **121 query distinte**.

Il primo risultato arriva prima della misura

Le frasi vere fallivano il doppio di quelle che avevamo costruito noi:

ORIGINE DELLE QUERY	QUANTE NON TROVANO NIENTE
costruite a tavolino	18%
raccolte da Google	38%

Il motivo e' istruttivo. Le nostre erano scritte con le parole del catalogo. Quelle vere sono scritte con le parole di chi lavora: **capicorda, faston, puntalini, a cricchetto, autofilettanti, testa svasata, sds max, orbitale, sverniciare**. Parole che il catalogo non contiene, e dove non c'e' la parola non c'e' il risultato.

Chi misura la propria ricerca con query che si e' inventato ottiene sempre un buon voto.

CASO STUDIO

La misura di partenza: la ricerca nativa di PrestaShop

PrestaShop costruisce un **indice invertito**, cioe' un elenco che, data una parola, dice subito in quali prodotti compare.

A ogni parola assegna un peso a seconda del campo da cui viene: il codice articolo pesa 10, il nome del prodotto 6, le etichette 4, la categoria e la marca 3, le caratteristiche 2, le descrizioni 1.

Quando arriva una domanda la spezza in parole, scarta quelle sotto i tre caratteri, cerca ogni parola **per prefisso** (chi cerca "nastr" trova "nastro"), e tiene solo i prodotti che contengono tutte le parole. Poi ordina per somma dei pesi.

Ha anche un recupero degli errori di battitura: se una parola non trova nulla, cerca nell'indice quella piu' somigliante. Funziona, ma costa.

query senza nessun risultato	38 su 121
tempo mediano	20 ms
tempo nel caso peggiore	2.292 ms

Il tempo peggiore merita una riga: **le risposte peggiori sono anche le piu' lente**. Quando il motore non trova la parola si mette a cercare quella somigliante in tutto il vocabolario, ci mette due secondi, e alla fine non trova comunque. Il cliente paga l'attesa piu' lunga per la risposta peggiore.

CASO STUDIO

Il passaggio a Meilisearch

Meilisearch e' un motore di ricerca che si affianca al negozio: il catalogo gli viene consegnato sotto forma di documenti, uno per prodotto, e lui li indicizza.

Le gradazioni: la ricerca ibrida

Meilisearch puo' lavorare in due modi contemporaneamente, cercando le parole e cercando il significato, e **un solo parametro decide quanto pesa l'uno e quanto l'altro**. Si chiama `semanticRatio` ed e' un numero da 0 a 1.

VALORE	COMPORTAMENTO
0	conta solo la ricerca per parole
0,3	tre decimi del punteggio dal significato, sette dalle parole
0,5	meta' e meta'
0,7	sette decimi dal significato, tre dalle parole
1	conta solo il significato, le parole vengono ignorate

Non e' una manopola del volume: e' una ripartizione, e le due parti devono fare uno. Non esistono valori sopra 1, e il motore li rifiuta.

Che cosa abbiamo mandato al modello degli embedding

Questa e' la decisione che incide di piu' e che nessuno vede. Al modello non va l'intero documento: va **un campo solo**, composto da noi, che contiene nell'ordine: titolo, marca, percorso del reparto, nomi popolari dell'oggetto, descrizione breve, descrizione lunga, caratteristiche tecniche in forma leggibile.

Restano fuori i codici, il prezzo, la giacenza, gli indirizzi delle immagini e i dati di imballaggio. Un codice articolo non ha un significato da rappresentare: sarebbe rumore che sposta il vettore senza aggiungere senso. Quei campi restano nel documento, cercabili per parole e utilizzabili nei filtri, semplicemente non entrano nel vettore.

C'e' anche un limite tecnico da conoscere: Meilisearch tronca il testo destinato al modello a una soglia configurabile. Con il valore che avevamo scelto all'inizio, **un prodotto su cinque veniva tagliato**, e la parte tagliata erano le caratteristiche tecniche, che stanno in fondo. Sarebbe stato un difetto invisibile: l'indice si costruisce, tutto sembra a posto, e un quinto del catalogo ha un vettore mutilato. Alzata la soglia, nessun prodotto viene piu' troncato.

Costi e tempi di costruzione

documenti caricati	1.535
tempo di caricamento, vettori compresi	15 secondi
costo degli embedding	circa 1,5 centesimi

CASO STUDIO

I sinonimi: una verifica che cambia l'approccio

Davanti a "flessibile" che non trova le smerigliatrici, la risposta ovvia e': mettiamo un sinonimo nel motore. Tutti i motori di ricerca hanno una tabella dei sinonimi, dove si dichiara che due parole vanno considerate equivalenti.

Funziona, ma non su tutti i tipi di ricerca. Lo abbiamo verificato con una prova costruita in modo che l'unica cosa a cambiare fosse la presenza del sinonimo: un indice separato con sei prodotti, tre smerigliatrici angolari e tre trapani, interrogato quattro volte con la stessa parola.

TIPO DI RICERCA	SINONIMO	PRIMO RISULTATO	PUNTEGGIO
solo parole	assente	**nessun risultato**	
solo parole	attivo	Smerigliatrice angolare	
solo significato	assente	Smerigliatrice angolare AEG	0,0157
solo significato	attivo	Smerigliatrice angolare AEG	0,0157

Le prime due righe dicono che **il sinonimo funziona**: sulla ricerca a parole, senza di lui la parola "flessibile" non trova niente, con lui escono le smerigliatrici.

Le ultime due dicono che **sulla ricerca per significato non fa assolutamente nulla**: con e senza il sinonimo il risultato e' identico, e identico fino al punteggio. Quel punteggio e' un numero calcolato dal motore: se il sinonimo avesse avuto un qualunque effetto, sarebbe cambiato. Non e' cambiato di un decimillesimo.

Che poi in quelle due righe esca comunque una smerigliatrice non e' merito del sinonimo: dipende dal fatto che in un indice di soli sei prodotti il vettore di "flessibile" finisce comunque piu' vicino a una smerigliatrice che a un trapano. Sul catalogo intero, millecinquecento prodotti, la ricerca per solo significato interrogata con "flessibile" le smerigliatrici non le trova affatto.

Perche' succede

Il vettore del prodotto viene calcolato **quando si carica il catalogo**, a partire dal testo della scheda. Nessuna tabella di sostituzione puo' cambiare un numero gia' scritto nell'indice. E la domanda del cliente viene trasformata in vettore cosi' com'e', senza passare dai sinonimi.

Quindi i sinonimi del motore funzionano al cento per cento quando si cerca per sole parole, a meta' nella ricerca mista, e per niente nella ricerca puramente semantica.

La strada alternativa, e perche' e' migliore

Se invece la parola la si scrive **dentro il catalogo**, funziona su tutti e due i canali. E ne guadagna un terzo.

DOVE SI METTE LA PAROLA	RICERCA A PAROLE	RICERCA PER SIGNIFICATO	GOOGLE
tabella dei sinonimi del motore	si'	no	no
dentro la scheda prodotto	si'	si'	si'

La tabella dei sinonimi la legge solo il motore di ricerca interno. La parola scritta nella scheda la legge anche Google. Un prodotto descritto come "dispositivo a corrente residua" non si trova cercando "differenziale" ne' dentro il sito ne' sui motori esterni.

La parola scritta nel catalogo si paga una volta e lavora su due canali: dentro il motore di ricerca del negozio e fuori, sui motori di ricerca esterni. Per questo la preferiamo alla tabella dei sinonimi, che resta uno strumento del motore e va usata per quello che sa fare.

I nomi del cliente

Da qui la soluzione adottata. Per ciascuna delle 75 categorie del catalogo abbiamo definito i nomi con cui la gente chiede quell'oggetto, che non sono quelli con cui lo descrive il produttore:

CATEGORIA A CATALOGO	NOMI DEL CLIENTE
Smerigliatrici angolari	flessibile, mola, mola a disco, frullino
Martelli perforatori	tassellatore, perforatore, martello demolitore
Pinze spellacavi	spellafili, spelafili, spelacavi
Chiavi esagonali	brugola, chiavi a brugola, chiave a L
Pistole pinzatrici manuali	sparapunti, graffettatrice, pinzatrice
Cacciaviti elettrici	avvitatore, giravite elettrico
Tagliacavi	tronchese, tronchesina, taglia cavi

Il dato vive **dentro il negozio**, non in un file esterno: un modulo aggiunge alla categoria un campo dedicato, che il cliente del sito non vede e che il gestionale del negozio puo' modificare. L'esportatore lo rilegge da li' e lo porta sia nel documento sia nel testo da cui si calcola il vettore.

E' la parte del lavoro che si ripaga meglio, perche' vale anche per il posizionamento sui motori di ricerca esterni.

Su PrestaShop il campo si aggiunge con un modulo; su WooCommerce con un campo personalizzato; su Magento con un attributo; su un e-commerce su misura con una colonna. Il principio non cambia: **un attributo della categoria, dentro il negozio, non esposto al cliente, esportabile.**

CASO STUDIO

Il risultato

Le stesse 121 query, lo stesso catalogo, gli stessi nomi del cliente dati a entrambi i motori, ciascuno nel campo che il suo programma sa leggere.

	PRESTASHOP NATIVO	MEILISEARCH A 0,3
query senza nessun risultato	36	0
tempo mediano	33 ms	112 ms
tempo nel caso peggiore	3.237 ms	168 ms

Il tempo mediano peggiora di novanta millesimi, e quei novanta millesimi comprendono il viaggio fino al centro dati e ritorno, mentre PrestaShop gira sulla stessa macchina del negozio.

In cambio **sparisce la coda lunga**: il caso peggiore passa da tre secondi e due a centosessantotto millesimi. Per chi usa un sito conta piu' il tetto che la media.

Perche' 0,3 e non un altro valore

Abbiamo misurato tutte e 121 le query a sei valori diversi.

VALORE	QUERY SENZA RISPOSTA	I CINQUE CODICI DI PROVA	TEMPO MASSIMO
0 (solo parole)	5	5 su 5	442 ms
0,2	0	5 su 5	142 ms
0,3	0	5 su 5	168 ms
0,5	0	5 su 5	156 ms
0,7	0	5 su 5	150 ms
1 (solo significato)	0	**1 su 5**	184 ms

Basta un venti per cento di significato per chiudere tutti i buchi. I codici articolo reggono fino a 0,7 e crollano solo all'estremo.

Guardando quanto cambiano i primi risultati passando da un valore al successivo si scoprono tre zone, non una scala continua:

PASSAGGIO	QUERY IN CUI CAMBIA LA TESTA DELLA CLASSIFICA
da 0 a 0,2	10 su 121
da 0,2 a 0,3	2
da 0,3 a 0,5	20
da 0,5 a 0,7	**92**
da 0,7 a 1	18

Fino a 0,5 il motore resta sostanzialmente lessicale e il significato tappa i buchi. Sopra 0,5 comanda il significato. **0,3 e' il valore prudente**: chiude tutti i buchi, non tocca i codici, e non stravolge il comportamento che il cliente del negozio si aspetta.

Qualche esempio concreto

QUERY	PRESTASHOP	MEILISEARCH A 0,3
`flessibile`	Estrattore di polvere Rowenta	Smerigliatrice angolare AEG WS 12-125
`tassellatore`	nessun risultato	Martello perforatore AEG BBH 12
`attrezzo per spelare i fili`	nessun risultato	Pinza spellacavi Microconnect PP-LSA2
`cosa serve per forare il muro`	nessun risultato	Martello perforatore AEG BH 24E SDS-plus
`vorrei un avvitatore leggero da usare in casa`	nessun risultato	Avvitatore elettrico AEG S 4000 E

QUERY	PRESTASHOP	MEILISEARCH A 0,3
per lucidare il pavimento	Lucidatrice Hoover F 3870	Lucidatrice per pavimenti AEG PE 150
BBH18-LI402C	Martello perforatore AEG BBH18-LI402C	Martello perforatore AEG BBH18-LI402C

Perche' i dati da soli non bastano

La ricostruzione del catalogo migliora qualcosa anche sul motore vecchio, e su singole parole si vede.

DOMANDA	MOTORE VECCHIO, PRIMA	MOTORE VECCHIO, DOPO
tassellatore	nessun risultato	30 risultati
flessibile	3 aspirapolveri	18 smerigliatrici
cacciavite	62 risultati	114 risultati

Ma sul totale delle centoventuno domande quel guadagno vale **due domande recuperate**: da trentotto senza risposta a trentasei. Tutto il resto, le altre trentasei, lo recupera il motore nuovo.

La ragione e' strutturale e non si aggira scrivendo meglio le schede. Un motore che confronta lettere non puo' rispondere a "cosa serve per forare il muro", perche' quelle quattro parole non compariranno mai in nessuna scheda, per quanto bene la si scriva. Le domande fatte a parole proprie, che su questo catalogo erano circa un terzo del totale, richiedono un motore che lavori sul significato.

I dati rimessi in ordine sono cio' che permette al motore nuovo di rendere al massimo. Il motore nuovo e' cio' che trasforma quel lavoro in risposte.

CASO STUDIO

Che cosa questi numeri non dicono

Va detto con chiarezza, perche' e' il limite della misura e chiunque sappia leggere un benchmark lo chiederebbe.

Abbiamo misurato che il motore **risponde**, non che risponda **bene**. I primi quattro risultati dei due motori coincidono su cinque query su centoventuno: abbiamo cambiato quasi tutta la classifica, e per stabilire che la classifica nuova sia migliore serve l'occhio di chi conosce la merce, su un campione.

Il dato solido e' questo: prima trentasei domande su centoventuno non ricevevano nessuna risposta, adesso la ricevono tutte, e il tempo di attesa nel caso peggiore e' diciannove volte piu' basso.

CASO STUDIO

In sintesi

I tre problemi che avevamo trovato all'inizio, e la cura di ciascuno:

PROBLEMA	ESEMPIO	CURA
il motore e' vecchio	`cacciavite` non trova i cacciaviti perche' cerca per prefisso	cambiare motore
i dati sono scritti male	il titolo dice `AEG BBH 12` e non dice che e' un martello	riscrivere il catalogo
il cliente descrive il problema	`cosa serve per forare il muro`	ricerca per significato

Sono tre lavori diversi, con tre costi diversi, e si possono affrontare anche uno per volta. Ma vanno fatti tutti e tre: la ricostruzione dei dati rende trovabile quello che c'e', il motore nuovo risponde a chi non sa come si chiama, e insieme portano le domande senza risposta da trentasei a zero.

Il significato chiude i buchi.
Le parole mantengono la precisione.

36 → 0	19×	0,3
QUERY SENZA RISPOSTA	TEMPO PEGGIORE PIÙ BASSO	RAPPORTO SEMANTICO SCELTO

MI

Massimo Ivaldi

A SEARCH COMPANY

IN SINTESI

Trovare di più.
Sbagliare di meno.

La ricerca semantica funziona quando catalogo, modello e misura vengono progettati insieme. Il significato chiude i buchi; le parole mantengono precisione su codici e nomi tecnici.

01 / Ricostruire

Scrivere il catalogo nella lingua del cliente, non soltanto in quella del produttore.

02 / Misurare

Usare query reali e un corpo di prova stabile, non esempi scelti per funzionare.

03 / Bilanciare

Mescolare parole e significato nella proporzione adatta a quel catalogo.

04 / Verificare

Controllare errori, codici, tempi peggiori e qualità della classifica.

massimoivaldi.com · A search company

settembre 2026