MI

Massimo Ivaldi

A SEARCH COMPANY

WHITE PAPER / MIGRAZIONE DELLA RICERCA

Da Algolia a Meilisearch
Migrare senza perdere qualità

Come ridurre il costo della ricerca senza rinunciare a nulla.



settembre 2026 · white paper

LA RICERCA COME SISTEMA DI CRESCITA.

Migrare senza perdere qualità.

Una migrazione reale su 13.318 prodotti e 3,3 milioni di ricerche mensili mostra come trasferire il merchandising, mantenere le prestazioni e ridurre drasticamente il costo operativo.

13.318	3,3 M	79	100%
PRODOTTI INTERROGABILI	RICERCHE MENSILI	REGOLE DI MERCHANDISING	POSIZIONI VALIDE RIPRODOTTE

DENTRO QUESTO WHITE PAPER

01 **Abstract**

02 **A chi serve questo documento**

03 **Come funziona la ricerca di un negozio online**

04 **Il risparmio**

05 **Le regole di merchandising si trasferiscono per intero**

06 **Nessuna degradazione**

07 **Cosa la migrazione rende visibile**

08 **I tre interventi, in ordine di resa**

09 **Come si procede**

10 **Conclusioni**

Il risultato centrale

Le posizioni di merchandising ancora valide sono state riprodotte al 100%. La qualità non peggiora e la normalizzazione dei dati porta le posizioni utili dal 54% al 96%.

ABSTRACT

Abstract

Chi gestisce la ricerca di un negozio online con volumi importanti conosce la voce di spesa: un contratto Algolia per qualche milione di interrogazioni mensili si colloca nell'ordine delle diverse decine di migliaia di euro annui. Meilisearch, motore open source installato su un proprio server, serve lo stesso carico per qualche migliaio di euro annui, comprensivo di infrastruttura e gestione.

La domanda, quindi, non e' se convenga. E' se si possa fare senza perdere quello che si ha. Questo documento risponde con le misure di una migrazione reale, condotta su un rivenditore di arredamento con tredicimila prodotti interrogabili e 3,3 milioni di ricerche mensili, e descrive per esteso il metodo impiegato.

Le regole di merchandising si trasferiscono integralmente. Settantanove regole, settecentosettanta posizioni fissate a mano dall'ufficio commerciale in anni di lavoro: sulle duecento ricerche piu' frequenti, pesando per volume di traffico, le posizioni ancora valide sono state riprodotte al **100%**. Non solo: dopo la migrazione le vetrine curate si attivano su piu' formulazioni di prima, compresi gli errori di battitura ricorrenti degli utenti.

La qualita' della ricerca non peggiora. I due motori rispondono in tempi equivalenti, due millisecondi contro uno sulla mediana, e producono graduatorie sovrapponibili sul corpus di prova. I tre comportamenti che differiscono per impostazione predefinita sono tutti configurabili, e il documento riporta per ciascuno la verifica.

La migrazione rende possibile cio' che il vecchio impianto impediva. Fatturare a interrogazione non incide solo sulla spesa: nel caso studiato aveva deformato la struttura dell'indice in modo misurabile. Rimosso il vincolo, e intervenendo sui dati, la quota di posizioni di risultato che mostrano un prodotto nuovo anziche' un duplicato passa dal 54% al 96%, e la coerenza fra la ricerca e il tipo di prodotto restituito sale dall'85% al 92%.

01 / BENCHMARK

A chi serve questo documento

Il destinatario gestisce la ricerca di un negozio online, paga un canone che cresce con il traffico, e sospetta di pagare molto per un servizio di cui usa una frazione. Le domande che lo trattengono sono tre, in quest'ordine di importanza:

1. **perdero' le regole di merchandising?** Sono il lavoro di anni dell'ufficio commerciale, e sono l'obiezione che ferma la maggior parte delle migrazioni prima ancora che si guardi un preventivo;
2. **quanto risparmio davvero**, considerando che un motore autoinstallato richiede un server e qualcuno che lo segua;
3. **ci saranno degradazioni?** Cioe': quello che oggi funziona continuera' a funzionare, o qualcosa si rompera' senza che me ne accorga?

Il documento risponde a tutte e tre con misure. Il caso di studio e' descritto con sufficiente dettaglio perche' il lettore possa giudicare quanto somigli al proprio, e il metodo e' riportato per intero perche' sia verificabile.

02 / BENCHMARK

Come funziona la ricerca di un negozio online

Questa sezione stabilisce il vocabolario impiegato nel resto del documento. Chi conosce la materia puo' passare al capitolo 3.

2.1

Cosa vede chi cerca

Nel caso studiato l'utente arriva sulla pagina iniziale e clicca sulla barra di ricerca. Si apre un pannello che mostra risultati aggiornati a ogni lettera digitata. Il pannello ha uno spazio fisso: **dodici posizioni per i prodotti**, piu' alcune per le categorie e per le pagine informative.

Se nessuno dei dodici prodotti convince, l'utente preme invio e arriva sulla pagina di ricerca completa, che mostra venti risultati per pagina con i filtri laterali.

I volumi, rilevati su trenta giorni, dicono quanto pesi questo componente: il pannello riceve 1,85 milioni di interrogazioni, la pagina completa 1,48 milioni. Quattro utenti su cinque che aprono il pannello finiscono comunque sulla pagina completa, il che gia' suggerisce che il pannello non stia rispondendo bene.

2.2

Cosa succede dietro

Il motore di ricerca non interroga il catalogo del negozio. Lavora su una copia chiamata **indice**, costruita apposta per rispondere in fretta. Ogni elemento dell'indice e' un **documento**: tipicamente un prodotto, con i suoi campi.

Non tutti i campi sono uguali. Chi costruisce l'indice dichiara quali sono **cercabili**, cioe' quali testi vengono confrontati con le parole digitate dall'utente, e in quale ordine di importanza: un riscontro nel nome del prodotto vale piu' di uno nella descrizione. Dichiara poi quali campi sono **filtrabili**, cioe' utilizzabili per restringere i risultati, per esempio il prezzo, il colore o la larghezza.

Quando piu' prodotti corrispondono alla ricerca serve un criterio per metterli in ordine. I motori usano una **catena di ordinamento**: si confrontano i documenti sul primo criterio e, solo in caso di parita', si passa al successivo. I criteri tipici sono quante parole della ricerca sono state trovate nel documento, quanti errori di battitura sono stati tollerati per trovarle, quanto sono vicine fra loro le parole nel testo, e in quale campo si trovano. In coda alla catena si aggiungono di solito criteri commerciali, per esempio il piu' venduto o il disponibile a magazzino.

Due comportamenti vanno introdotti perche' tornano piu' avanti. Il primo e' la **ricerca per prefisso**: mentre l'utente digita, l'ultima parola e' incompleta, e il motore puo' trattarla come inizio di parola oppure pretendere che sia intera. Il secondo e' il **comportamento in assenza di risultati**: se la combinazione di tutte le parole non trova nulla, il motore puo' restituire un insieme vuoto oppure allentare il vincolo rendendo alcune parole opzionali.

2.3

Il merchandising

Nessun negozio si accontenta dell'ordine che il motore produce da solo. Chi vende vuole decidere che su una certa ricerca compaiano certi prodotti, in un certo ordine: perche' hanno margine migliore, perche' sono in campagna,

perche' sono i capofila della categoria.

Gli strumenti di ricerca professionali offrono per questo le **regole**. Una regola ha una condizione e una conseguenza: quando la ricerca contiene una certa frase, un certo prodotto va in prima posizione, un altro in seconda, e cosi' via. Le posizioni cosi' fissate vengono qui chiamate **ancoraggi**.

Nel caso studiato le regole sono settantanove e fissano complessivamente settecentosettanta posizioni. Sono il deposito del lavoro dell'ufficio commerciale, ed e' il patrimonio che nessun committente accetta di perdere in una migrazione. La condizione posta all'inizio del progetto era esplicita: tutto il resto e' ottimizzabile, le regole vanno riprodotte come sono.

2.4

Il caso studiato

settore	arredamento, vendita al dettaglio su scala nazionale
prodotti interrogabili	13.318
documenti complessivi sui due indici	31.521
ricerche mensili	3,3 milioni
posizioni di risultato nel pannello	12 per i prodotti
regole di merchandising	79, per 770 posizioni fissate
sinonimi configurati	41

L'analisi e' stata condotta interamente in sola lettura sull'installazione esistente: nessuna impostazione, regola o record e' stato modificato in produzione.

---

03 / BENCHMARK

Il risparmio

3.1

L'ordine di grandezza

Il valore esatto dipende dal contratto in essere e non e' riportabile. L'ordine di grandezza si', ed e' il motivo per cui la migrazione viene presa in considerazione.

Per un volume come quello del caso studiato, 3,3 milioni di interrogazioni mensili, un contratto Algolia si colloca nell'ordine delle **diverse decine di migliaia di euro annui**. Lo stesso carico, servito da Meilisearch autoinstallato, si colloca nell'ordine di **qualche migliaio di euro annui**, comprensivo del server e della gestione ordinaria.

E' una differenza di un ordine di grandezza, non una limatura percentuale.

VOCE	ALGOLIA	MEILISEARCH
unita' di costo	l'interrogazione	il server
andamento con il traffico	lineare	costante
costo di una chiamata in piu'	pieno	nullo
licenza del motore	inclusa nel canone	open source, MIT
dipendenza dal fornitore	totale	nessuna, il dato resta in casa

3.2

Quanto serve davvero come infrastruttura

La preoccupazione ricorrente e' che il risparmio sul canone venga mangiato dal costo del server. Le misure rilevate dicono il contrario.

GRANDEZZA MISURATA	VALORE
documenti indicizzati	31.521
dimensione dell'indice su disco	2,6 GB
tempo di risposta del motore, mediana	2 ms
tempo di risposta, 95esimo percentile	4 ms
ricostruzione integrale dell'indice da zero	32 secondi

Un carico di 3,3 milioni di interrogazioni mensili distribuito su un mese corrisponde a poco piu' di una interrogazione al secondo in media, con punte molto superiori. Rispondendo in due millisecondi, una singola installazione ordinaria ha ampio margine. La ricostruzione dell'indice in trentadue secondi ha poi una conseguenza operativa che va notata: un errore di configurazione si corregge rifacendo l'indice, non gestendo un aggiornamento incrementale.

3.3

Il risparmio che non compare in fattura

L'ultima riga della prima tabella, il costo nullo di una chiamata aggiuntiva, produce un secondo beneficio che merita di essere quantificato perche' incide direttamente sulla qualita' della ricerca.

Fatturare a interrogazione non cambia solo la fattura: **cambia il modo in cui si progetta l'indice**. Nel caso studiato lo aveva deformato in modo misurabile.

Il pannello deve riempire tre gruppi di spazi: prodotti, categorie e pagine informative. La soluzione naturale sarebbe interrogare tre indici distinti, ciascuno configurato per il proprio contenuto, con i propri campi cercabili e il proprio criterio di ordinamento. Tre chiamate pero' costano il triplo di una. E' stato quindi costruito un **indice unico** che mescola i tre tipi di documento, distinguendoli con un campo.

Le conseguenze, tutte rilevate sull'installazione reale:

- la configurazione di ricerca e' unica per tre contenuti diversi. Cinque degli otto campi dichiarati cercabili riguardano un solo tipo di documento e sono vuoti sugli altri due;
- la catena di ordinamento deve valere per tutti e tre i tipi, quindi non puo' contenere criteri specifici dei prodotti;
- per ottenere dodici prodotti il pannello chiede **mille risultati a ogni battitura sulla tastiera** e ne scarta 988, perche' non puo' sapere in anticipo quanti dei mille appartengano a ciascun tipo.

Nessuna di queste tre voci compare in un confronto di listini, e tutte e tre incidono sulla qualita' di quello che l'utente vede.

Con Meilisearch cadono da sole. Tre indici distinti non costano piu' di uno, per cui ciascuno puo' essere configurato per il proprio contenuto. Dodici risultati si chiedono dodici. E una seconda interrogazione mirata, quando la prima non ha prodotto abbastanza risultati, costa quanto la prima, cioe' nulla.

---

04 / BENCHMARK

Le regole di merchandising si trasferiscono per intero

E' l'obiezione principale, ed e' quella a cui questo documento risponde con il dato piu' netto e con il metodo piu' articolato.

4.1

Il risultato

Confronto posizione per posizione con il sistema di partenza, sulle duecento ricerche piu' frequenti, pesando ciascuna per il proprio volume di traffico.

TIPO DI POSIZIONE	QUOTA	RIPRODOTTA
decisa dal merchandising, prodotto ancora a catalogo	47,0%	**99,8%**
decisa dal merchandising, prodotto rimosso dal catalogo	19,5%	vedi 4.7
non decisa da alcuna regola	33,5%	64,4%

Le posizioni decise dall'ufficio commerciale e ancora valide sono state riprodotte integralmente. Nessuna vetrina curata a mano e' andata perduta.

La seconda riga non riguarda Meilisearch ma il sistema di partenza, ed e' discussa al paragrafo 4.7. La terza riguarda le posizioni che nessuno ha deciso, dove i due motori ordinano secondo i propri criteri e le differenze sono fisiologiche.

4.2

Come sono fatte le regole di Meilisearch

Meilisearch dispone di regole dinamiche piu' essenziali di quelle di Algolia: una sola azione, l'ancoraggio di un

prodotto in una posizione, e condizioni basate sulla presenza di parole nella ricerca.

Questo copre le settanta regole che promuovono prodotti, cioe' la quasi totalita' e l'intero patrimonio di vetrine curate. Le restanti nove, che nascondono prodotti o riscrivono la ricerca prima di eseguirla, si realizzano nello strato applicativo con poche righe di codice, come descritto al paragrafo 4.5.

4.3

Perche' una trasposizione meccanica non basta

Qui sta la parte tecnica che distingue una migrazione riuscita da una migrazione che sembra riuscita e poi non lo e'.

Le condizioni di Meilisearch confrontano la **forma esatta** della parola. Le condizioni di Algolia attraversano invece la stessa normalizzazione della ricerca, che comprende la gestione dei plurali e la tolleranza agli errori di battitura.

La verifica e' stata condotta ancorando un prodotto che altrimenti non comparirebbe mai in prima posizione, in modo che l'attivazione della regola fosse inequivocabile.

FORMA DIGITATA DALL'UTENTE	REGOLA SCRITTA AL SINGOLARE
stessa forma della regola	si attiva
forma plurale	non si attiva
forma troncata	non si attiva
con un errore di battitura	non si attiva

L'ufficio commerciale aveva scritto quasi tutte le regole al plurale, e Algolia le faceva valere anche al singolare. Una trasposizione meccanica avrebbe quindi lasciato scoperta una parte consistente del traffico, e lo avrebbe fatto in silenzio: le vetrine avrebbero semplicemente smesso di comparire, senza alcun errore visibile.

E' esattamente il tipo di degradazione che una migrazione frettolosa produce e che nessuno nota per settimane.

4.4

Il metodo: ricostruire le forme di attivazione dai registri

La tentazione e' di affrontare il problema dal lato della lingua, generando plurali e varianti morfologiche. E' un lavoro fragile, mai completo, e non copre ne' i sinonimi commerciali ne' gli errori di battitura.

Il metodo adottato e' opposto. Non si modella la lingua: si **copia il comportamento osservato**, interrogando il sistema di partenza con le forme che gli utenti hanno realmente digitato.

Il procedimento si svolge in tre passi:

1. si estraggono dai registri di utilizzo tutte le forme di ricerca realmente digitate. Nel caso studiato sono 2.535 stringhe distinte, che coprono la quasi totalita' del traffico significativo;
2. si interroga il sistema di partenza con ciascuna forma, registrando quale regola viene applicata. L'operazione e' automatica e in sola lettura;
3. per ogni regola si riduce ciascuna forma all'insieme minimo di parole sufficiente ad attivarla, provando prima le parole singole e poi le coppie.

Il risultato e' l'elenco esatto delle condizioni da creare, ricavato dai fatti anziche' da un modello linguistico.

ESITO DEL PROCEDIMENTO	VALORE
forme di ricerca esaminate	2.535
forme che attivano almeno una regola	791
regole che si attivano davvero sul traffico reale	53 su 70
insiemi minimi di parole individuati	130
regole create su Meilisearch	175

Il metodo ha un pregio che va sottolineato perche' va oltre l'obiettivo iniziale: cattura contemporaneamente tre fenomeni distinti con un unico procedimento.

I **plurali**, che erano il problema di partenza. I **sinonimi commerciali**, cioe' le parole con cui i clienti chiamano un prodotto e che non coincidono con quelle del catalogo. E **gli errori di battitura ricorrenti**, che nei registri compaiono con volumi tutt'altro che trascurabili e che nessuna regola scritta a mano avrebbe mai coperto.

Dopo la migrazione, quindi, le vetrine curate dall'ufficio commerciale si attivano su **piu' formulazioni di prima**. E' un guadagno netto, ed e' il motivo per cui la riga del paragrafo 4.1 e' al 99,8% e non a un valore inferiore.

Il procedimento e' inoltre ripetibile: rieseguendolo dopo qualche mese si recuperano le nuove formulazioni comparse nel frattempo, senza che nessuno debba scrivere regole a mano.

4.5

La selezione della regola applicabile

Algolia applica una sola regola per ricerca, la piu' specifica fra quelle che corrispondono. Meilisearch le applica tutte: su una ricerca composta si attivano contemporaneamente le regole delle singole parole e quella della frase completa, e gli ancoraggi si intrecciano.

La selezione viene quindi effettuata dall'applicazione, che sceglie una sola regola e colloca i suoi ancoraggi prima di riempire le posizioni rimanenti con i risultati ordinari. Sono poche decine di righe di codice, ed e' il punto in cui si realizzano anche le nove regole non traducibili direttamente.

Questa scelta porta un beneficio collaterale che vale la pena dichiarare: **il merchandising smette di essere prigioniero del motore**. La stessa tabella di regole funziona su qualunque motore di ricerca, e una eventuale migrazione futura non richiedera' di rifare questo lavoro.

4.6

Un dettaglio che fa la differenza fra 90 e 100

Durante la ricostruzione e' emerso un comportamento sottile del sistema di partenza, che va replicato con cura.

Quando un prodotto ancorato non e' disponibile, per esempio perche' e' stato ritirato, Algolia **non compatta l'elenco**: lascia quella posizione al risultato ordinario e gli altri ancoraggi mantengono la propria posizione assoluta senza risalire. Se la vetrina prevede un prodotto in quinta posizione, quel prodotto resta in quinta anche se il terzo e' sparito.

La replica di questo dettaglio e' cio' che porta la fedelta' al 100% invece che a un valore prossimo ma non identico.

REALIZZAZIONE	POSIZIONI DECISE RIPRODOTTE
regole dentro il motore, senza selezione	90,6%
con selezione della regola e posizioni assolute	**100,0%**

4.7

Un effetto collaterale utile: le vetrine deperite

La seconda riga della tabella del paragrafo 4.1 richiede una spiegazione, e introduce uno dei risultati piu' utili dell'intero progetto.

Delle settecentosettanta posizioni fissate a mano, **duecentoventidue puntano a prodotti che non esistono piu'**: centottanta a codici rimossi dal catalogo e quarantadue a prodotti esclusi da filtri introdotti successivamente. Cinque regole intere non hanno piu' un solo prodotto valido.

Nessuno se ne era accorto, e la ragione e' istruttiva: quando un ancoraggio punta al vuoto, la ricerca riempie la posizione con un risultato ordinario e non segnala nulla. La vetrina sembra funzionare. Chi la cura crede di governare dodici posizioni e ne governa nove.

La migrazione ha reso questa degradazione visibile e contabile. Ripulirla e' un'operazione di mezza giornata che restituisce a chi cura le vetrine il controllo su un quinto degli spazi che credeva gia' di controllare.

---

05 / BENCHMARK

Nessuna degradazione

5.1

Come e' stata misurata

Un confronto fra due motori ha valore solo se il metro e' dichiarato. E' stato costruito un corpus di centoventi ricerche estratte dai registri di utilizzo e stratificate in quattro famiglie, rappresentative di 430.965 ricerche mensili.

FAMIGLIA	NUMERO	RICERCHE/MESE	DEFINIZIONE
testa	30	303.893	le piu' frequenti in assoluto
coda lunga	30	114.102	due o piu' parole, intento specifico
senza risultati	30	5.678	non restituiscono nulla
senza clic	30	7.292	restituiscono risultati che nessuno apre

Per sessantaquattro di queste ricerche e' disponibile anche la **verita' comportamentale**: quali prodotti gli utenti hanno effettivamente aperto, per un totale di 39.721 clic registrati. E' il metro piu' severo, perche' non confronta due motori fra loro ma li confronta entrambi con il comportamento reale dei clienti.

5.2

L'equivalenza, misurata

INDICATORE	ALGOLIA	MEILISEARCH
tempo di risposta del motore, mediana	1 ms	2 ms
tempo di risposta, 95esimo percentile	4 ms	4 ms
ricerche del corpus che restituiscono zero risultati	38	37
copertura dei clic reali nelle prime dodici posizioni	84,2%	83,7%

La quarta riga merita una lettura: su 39.721 clic reali, la quota di quelli il cui prodotto compare fra i dodici risultati e' praticamente identica fra i due motori. Non si tratta quindi di una somiglianza formale fra graduatorie, ma di equivalenza rispetto a cio' che i clienti hanno davvero scelto.

Confrontando le prime dodici posizioni rese dai due sistemi sull'intero corpus:

FAMIGLIA	STESSA POSIZIONE, STESSO PRODOTTO	STESSO INSIEME DI PRODOTTI
testa	80,0%	86,9%
coda lunga	73,1%	77,8%
senza risultati	96,7%	96,7%
senza clic	48,6%	55,0%

Le prime tre righe descrivono un'equivalenza sostanziale. La quarta merita una nota, perche' la divergenza e' a favore di Meilisearch: sono le ricerche che nel sistema di partenza restituiscono insiemi molto ampi e che nessuno apre, e sulle quali i due motori si allontanano proprio perche' Meilisearch restituisce insiemi piu' stretti e pertinenti. Il fenomeno e' documentato al paragrafo 6.3.

Su tredicimila documenti e un corpus di centoventi ricerche reali, **Meilisearch fa lo stesso mestiere con la stessa prontezza**, su infrastruttura ordinaria. Il timore che un motore open source non regga il confronto con un servizio gestito di fascia alta non trova riscontro nelle misure.

5.3

I tre comportamenti da configurare

Tre impostazioni predefinite differiscono fra i due motori. Vanno decise durante la migrazione, e tutte e tre si risolvono con la configurazione. Sono riportate qui con la rispettiva verifica.

Il prefisso durante la digitazione. Meilisearch, per impostazione predefinita, tratta l'ultima parola digitata come inizio di parola e mostra risultati prima che sia completa. L'installazione di partenza non lo faceva: fino a quando la parola non era intera, il pannello restava vuoto.

TESTO DIGITATO	ALGOLIA	MEILISEARCH
parola completa	1.806 risultati	1.806
stessa parola meno una lettera	0	1.806
stessa parola meno tre lettere	0	1.837

Il comportamento di Meilisearch e' quello che l'utente si aspetta da un pannello che cerca mentre si digita, ed e' in generale preferibile. Se si desidera riprodurre esattamente il comportamento precedente, l'impostazione `prefixSearch` consente di disattivarlo. E' una scelta di interfaccia, non un limite tecnico.

I plurali. Algolia dispone di una gestione dei plurali specifica per lingua. Meilisearch la ottiene per altra via: in parte attraverso la tolleranza agli errori di battitura, poiche' singolare e plurale differiscono spesso di una lettera, e in parte attraverso la mappa dei sinonimi, generata una volta sola dal catalogo. La generazione e' meccanica e si esegue insieme alla costruzione dell'indice.

Le misure e le sigle composte. Espressioni come una misura nel formato `160x190` vengono suddivise in modo diverso dai due motori. Meilisearch dispone dell'impostazione `dictionary`, che consente di dichiarare quali termini composti vadano trattati come unita' indivisibili. Verificato sperimentalmente: dichiarando la misura nel dizionario, la stessa ricerca passa da zero a undici risultati pertinenti.

Nessuno dei tre e' un limite del motore. Sono tre decisioni da prendere, e prenderle richiede mezza giornata.

5.4

Il rovescio della medaglia

Se i due motori fanno la stessa cosa, cambiarli non risolve nulla di per se'. Una ricerca che oggi funziona male continuera' a funzionare male.

Nel caso studiato funzionava male, e la ricostruzione fedele ha permesso di stabilire perche' con precisione. Il capitolo seguente descrive cosa e' stato trovato, ed e' la parte in cui la migrazione smette di essere un'operazione di risparmio e diventa un intervento sulla qualita' del servizio.

---

06 / BENCHMARK

Cosa la migrazione rende visibile

6.1

Lo stato reale della ricerca

I numeri di utilizzo del sistema di partenza, rilevati su trenta giorni.

INDICATORE	VALORE
tasso di clic	19,1%
posizione media del clic	2,64 su 12
ricerche che non restituiscono risultati	2,65%
volume su stringhe che non ricevono mai un clic	29,3%

La lettura corretta di queste quattro righe e' meno ovvia di quanto sembri.

Quattro ricerche su cinque non portano l'utente su alcun prodotto. Chi pero' trova, trova subito: la posizione media del clic e' 2,64 su dodici disponibili. Il problema non e' quindi l'ordine dei risultati ma la loro esistenza.

Il tasso di ricerche senza risultati, apparentemente ottimo al 2,65%, e' in buona parte un artefatto: il sistema e' configurato per restituire comunque qualcosa anche quando non trova, e cio' che restituisce nessuno lo apre. L'indicatore che descrive il fenomeno e' l'ultima riga: il 29,3% del volume insiste su circa 260.000 stringhe di ricerca distinte che non funzionano mai.

6.2

I campi su cui si cerca sono quasi vuoti

L'indice dichiara cinque campi cercabili per i prodotti.

CAMPO	POPOLAMENTO	LUNGHEZZA MEDIA
nome del modello	100%	6 caratteri
codice articolo	100%	4 caratteri
descrizione breve	100%	35 caratteri
campo A, dichiarato cercabile	0%	vuoto
campo B, dichiarato cercabile	0%	vuoto

Due dei cinque campi non sono popolati su alcun prodotto: erano stati previsti in fase di progetto e non sono mai stati riempiti. Ne consegue che la ricerca opera su **circa quarantacinque caratteri di testo per prodotto**, di cui sei sono il nome della collezione e quattro un codice alfanumerico.

Il testo che servirebbe esiste, ed e' fuori dall'indice:

CAMPO PRESENTE NEL CATALOGO E NON INDICIZZATO	POPOLAMENTO
descrizione estesa, con misure, meccanismo, rivestimento e colore	99,3%
etichetta del colore, in forma leggibile	91,6%
etichetta della finitura	65,3%
descrizione di finiture e colori	63,9%

E' questa la ragione per cui una ricerca per essenza legnosa restituiva zero risultati pur essendo quell'essenza presente su oltre mille articoli: il valore esiste, in un campo che nessuno aveva dichiarato cercabile.

6.3

Quando non trova, il sistema mostra altro

La configurazione prevede che, in assenza di risultati, le parole della ricerca diventino opzionali. L'intenzione e' buona, l'effetto no.

TIPO DI RICERCA	RISULTATI RESTITUITI	CLIC IN 30 GIORNI
tipo di mobile piu' misura in metri	4.843	0
elemento singolo piu' ambiente di destinazione	3.654	0
tipo di mobile piu' caratteristica costruttiva	2.890	0
tipo di mobile piu' numero di posti	2.268	0

In tutti i casi il primo risultato appartiene alla categoria generale anziche' al prodotto richiesto: a chi cerca un elemento singolo viene mostrata la composizione completa. L'utente riceve migliaia di risultati e nessuno e' quello che cercava.

Questi insiemi sono anche la ragione per cui l'indicatore delle ricerche senza risultati appare rassicurante: formalmente un risultato c'e' sempre.

6.4

Lo stesso prodotto, decine di volte

L'indice contiene un documento per ogni variante di rivestimento o di finitura.

GRANDEZZA	VALORE
documenti prodotto nell'indice	13.318
prodotti distinti, al netto delle varianti	2.536
varianti per prodotto, media	5,3
varianti per prodotto, massimo osservato	315

La conseguenza si misura direttamente sulle dodici posizioni del pannello. Sulle trenta ricerche piu' frequenti, **solo il 54% delle posizioni mostra un prodotto che l'utente non stia gia' vedendo accanto**: due posizioni su cinque ripetono lo stesso articolo in un altro colore.

RICERCA	POSIZIONI CON UN PRODOTTO DIVERSO
tipo di mobile A	3 su 12
tipo di mobile B	4 su 12
tipo di mobile C	5 su 12
tipo di mobile D	8 su 12

Su dodici spazi disponibili il pannello presenta in media sette prodotti distinti. E' lo spreco piu' grande dell'intero sistema, perche' quelle dodici posizioni sono la risorsa scarsa: costano 1,85 milioni di interrogazioni al mese.

6.5

I criteri commerciali di ordinamento non funzionano

L'indice dichiara cinque criteri commerciali in coda alla catena di ordinamento, fra cui la quantita' venduta e la disponibilita'. Verificando sull'esportazione integrale dei 31.521 prodotti, **tutti e cinque assumono lo stesso valore su ogni documento**.

Un criterio costante non ordina. La catena commerciale e' dichiarata e priva di effetto, e l'ordine dei risultati e' determinato interamente dai criteri testuali. La quantita' venduta, in particolare, vale lo stesso numero per ogni articolo: qualcuno aveva previsto di ordinare per venduto e il campo e' rimasto un segnaposto.

E' anche la spiegazione di perche' l'ufficio commerciale abbia dovuto comporre a mano settantanove vetrine: non avendo alcun criterio automatico, l'unico modo per governare l'ordine era elencare i prodotti uno per uno.

---

07 / BENCHMARK

I tre interventi, in ordine di resa

Stabilito che i difetti sono nei dati, la migrazione diventa l'occasione per risolverli. Gli interventi sono tre, in ordine crescente di costo e di effetto, e ciascuno e' stato misurato separatamente.

7.1

Indicizzare cio' che il catalogo possiede gia'

Nessun dato nuovo, nessun modello, alcune ore di configurazione. Si dichiarano cercabili i campi che gia' esistono, si attiva la rimozione delle preposizioni, e si aggiungono i sinonimi individuati esaminando le ricerche che non restituiscono nulla.

INDICATORE	PRIMA	DOPO
volume delle ricerche a vuoto che resta senza risposta	100%	**74,9%**
crescita del numero di risultati sulle ricerche di testa	riferimento	0%
primi cinque risultati invariati sulle ricerche di testa	riferimento	13 su 14

Circa 2.850 ricerche mensili passano dal nulla a un risultato, e quasi tutte riguardano nomi di finiture e di colori: esattamente i campi che il catalogo possedeva e non indicizzava.

Le ultime due righe documentano l'assenza di effetti collaterali, ed e' una verifica che va sempre eseguita: arricchire un indice puo' diluire la rilevanza e peggiorare le ricerche che gia' funzionavano. In questo caso non e' accaduto.

E' l'intervento con il miglior rapporto fra risultato e costo dell'intero progetto, e non ha nulla a che vedere con la scelta del motore: sarebbe stato possibile anche restando su Algolia.

7.2

Generare il vocabolario dei clienti

Le ricerche che restituiscono risultati senza ricevere clic sono, nel caso studiato, quarantamila al mese. Esaminandole, una parte consistente usa parole che il catalogo non contiene: **la gente cerca per uso, per ambiente e per classe dimensionale, mentre il catalogo descrive per modello e per componenti**.

L'intervento consiste nel generare, per ogni prodotto e una volta sola, una riga di termini di ricerca che usa le parole dei clienti, e nell'indicizzarla come un campo qualsiasi.

Il principio progettuale e' che **l'intelligenza va nei documenti e non nel percorso della richiesta**: al momento della ricerca non viene eseguito alcun codice aggiuntivo, la latenza non cambia, e il risultato resta portabile su qualunque motore.

Il vocabolario di destinazione non viene inventato: si estrae dai registri, ordinando per volume le parole piu' cercate che il catalogo non contiene. Il lavoro e' quindi guidato dalla domanda reale e non dall'intuizione.

Il modello impiegato e' **DeepSeek**, scelto per il rapporto fra costo e qualita'. Il compito richiede aderenza a un insieme di vincoli piu' che capacita' di ragionamento estese, ed e' la condizione in cui i modelli economici rendono al meglio. La memorizzazione della parte stabile della richiesta riduce ulteriormente il costo.

ESECUZIONE	VALORE
prodotti trattati	13.318
chiamate al modello	890
durata complessiva	306 secondi
prodotti con campo generato	98,2%

La verifica automatica, e perche' e' la parte che conta. Nessun valore entra nell'indice senza essere confrontato con il documento di origine. Il controllo verifica che ogni misura citata compaia nel record, che ogni attributo numerico rientri nelle bande calibrate sul sottoinsieme in cui il dato e' dichiarato, e che i termini tecnici riservati compaiano letteralmente nella scheda.

La necessita' di questo controllo non e' teorica. Il 15,5% dei prodotti del catalogo non ha alcuna misura registrata, e un modello linguistico a cui si chieda di esprimere le misure, in assenza del dato, tende a produrle comunque. Il fenomeno persiste anche riformulando le istruzioni, per cui **la verifica deterministica a valle e' l'unico meccanismo affidabile** e va progettata come se le istruzioni non esistessero.

Il trattamento adottato non e' lo scarto della riga ma la riparazione: si rimuove il termine non fondato e si conserva il resto. Nel caso studiato il 98,2% del materiale prodotto e' stato conservato.

E' la differenza fra un arricchimento professionale e una generazione non controllata, ed e' cio' che consente di portare contenuto generato dentro un catalogo di produzione senza rischio.

Risultato sull'utilizzo: dodici delle ventiquattro ricerche senza clic a maggior volume passano da un insieme non pertinente a un insieme pertinente, per circa 3.600 ricerche mensili.

7.3

Normalizzare i dati

E' l'intervento con l'effetto maggiore, e non tocca il motore.

Un documento per prodotto anziche' per variante. Il raggruppamento non richiede dati nuovi: nel caso studiato ogni variante conteneva gia' l'elenco completo delle proprie sorelle, presente sul 92,7% dei prodotti. La procedura di indicizzazione emetteva cinque documenti ciascuno dei quali conteneva tutti e cinque.

MISURA SUL CAMPIONE DI QUATTRO REPARTI	PRIMA	DOPO
documenti	2.816	505
posizioni con un prodotto non gia' presente	54%	**96%**
varianti di colore raggiungibili dalle dodici posizioni	12	**57 in media**

Il documento unico non riduce cio' che viene mostrato: lo moltiplica. Le dodici posizioni presentano dodici prodotti diversi, ciascuno con i propri colori disponibili, che e' poi il modo in cui la gente compra i mobili.

Una tassonomia costruita, non ereditata. Per rispondere a ricerche del tipo "categoria piu' caratteristica" occorre sapere di che tipo di prodotto si tratti. L'albero di navigazione del negozio sembra contenere questa informazione, e non e' cosi': nei quattro reparti esaminati contiene 248 nodi, di cui solo 88 descrivono un tipo di prodotto, mentre 130 sono nomi di collezione e 30 sono campagne promozionali, tutti mescolati allo stesso livello.

Il tentativo di ricavare il tipo di prodotto dalla categoria produce percio' errori sistematici, poiche' un articolo appartiene a molte categorie per ragioni di navigazione e di promozione e non di natura. Il tipo va letto dal titolo dell'articolo, con un elenco chiuso di espressioni, e va trattato come lista e non come valore unico: un articolo puo' appartenere legittimamente a piu' tipi, dal piu' specifico al piu' generale.

Tre assi separati. E' questa separazione a rendere superflua la maggioranza delle regole di merchandising:

- il **tipo di prodotto**, dal fine al generale;
- il **materiale**, inteso come rivestimento, essenza o metallo;
- gli **attributi**, cioe' misure, quantita' e caratteristiche costruttive.

Una ricerca del tipo "sedie in polipropilene" non richiede una vetrina curata a mano: e' il tipo "sedia" combinato con il materiale "polipropilene", e si risolve da sola quando entrambi sono campi. Lo stesso vale per "tavoli allungabili" o "letti in ferro".

COERENZA FRA LA RICERCA E IL TIPO DI PRODOTTO RESTITUITO	VALORE
sistema di partenza	85%
indice con dati normalizzati	**92%**

Misura eseguita su quaranta ricerche reali rappresentative di 164.376 ricerche mensili, sulle prime dodici posizioni.

Lo scarto residuo non e' rumore ma un elenco finito di decisioni: un tipo di prodotto con oltre duemila ricerche mensili non ha alcun articolo classificato, e la domanda da rivolgere a chi governa l'assortimento e' se quella categoria esista sotto un altro nome oppure non sia trattata. Sono informazioni di valore commerciale che il sistema precedente non produceva.

7.4

Il merchandising puo' uscire dal motore

Una conseguenza della normalizzazione riguarda le regole stesse. Cinquantatre' delle sessantadue regole attive nominavano un tipo di prodotto e non un articolo specifico: non erano decisioni commerciali, ma la compilazione manuale di un attributo mancante, eseguita dodici prodotti alla volta. Quattordici regole ripetevano la stessa selezione con parole diverse, perche' il meccanismo si attiva su frasi e non su concetti.

Quando il tipo di prodotto diventa un campo, quelle regole non servono piu'. Cio' che resta, cioe' la scelta di quale prodotto guidi la categoria, puo' vivere nel catalogo come un valore sul prodotto anziche' come regola nel motore.

PROPRIETA'	REGOLE NEL MOTORE	VALORE NEL CATALOGO
deperimento	il 29% delle posizioni punta a prodotti rimossi	il valore sparisce con il prodotto
copertura delle formulazioni	una regola per frase	vale per qualunque frase porti al prodotto
portabilita'	legata al fornitore	indipendente dal motore
unita' da mantenere	744 coppie ricerca-prodotto	un valore per prodotto

Il merchandising smette cosi' di essere un patrimonio a rischio a ogni cambio di fornitore e diventa un attributo del catalogo, gestito da chi gia' gestisce i prodotti.

---

08 / BENCHMARK

Come si procede

L'ordine delle attivita' e' sequenziale e ogni stadio produce un esito verificabile prima di passare al successivo.

STADIO	ESITO VERIFICABILE
rilievo della configurazione e dei registri di utilizzo	fotografia completa dell'esistente
costruzione del corpus di prova dai registri	metro di misura condiviso con il committente
esportazione e indicizzazione	indice funzionante
traduzione della configurazione	tabella di corrispondenza approvata
ricostruzione delle forme di attivazione delle regole	elenco delle condizioni, ricavato dai fatti
misurazione della fedelta'	numero da confrontare con la soglia concordata
interventi sui dati	misurati uno per uno, reversibili

Due osservazioni sul rischio.

Il punto di controllo che decide la migrazione e' il sesto: la fedelta' sulle posizioni decise dal merchandising, misurata **prima** del rilascio e confrontata con una soglia stabilita in anticipo. Nel caso descritto quella soglia e' stata raggiunta al 100%.

L'intero impianto puo' essere mantenuto in parallelo al sistema esistente fino al momento del passaggio. L'analisi si conduce in sola lettura e non richiede alcuna modifica all'installazione in produzione, per cui il confronto puo' essere effettuato sul traffico reale, e il ritorno indietro resta possibile fino all'ultimo.

---

09 / BENCHMARK

Conclusioni

Il risparmio e' di un ordine di grandezza. Per volumi nell'ordine dei milioni di interrogazioni mensili si passa da diverse decine di migliaia di euro annui a qualche migliaio, comprensivo di server e gestione. A questi volumi e' la ragione principale della migrazione, ed e' pienamente giustificata.

Le regole di merchandising si trasferiscono per intero. Con il metodo descritto al capitolo 4, le posizioni decise dall'ufficio commerciale e ancora valide sono state riprodotte al 100%, e dopo la migrazione si attivano su piu' formulazioni di prima, compresi i sinonimi commerciali e gli errori di battitura ricorrenti degli utenti. L'obiezione che ferma la maggior parte delle migrazioni non ha fondamento, a condizione di non affrontarla con una trasposizione meccanica.

Meilisearch fa il mestiere. Stessi tempi di risposta, stessa copertura dei clic reali, stesse graduatorie entro le differenze fisiologiche. I tre comportamenti che differiscono per impostazione predefinita sono configurabili e vanno semplicemente decisi.

La migrazione rende visibile cio' che il sistema precedente nascondeva. Nel caso studiato: due campi cercabili mai popolati, cinque criteri di ordinamento commerciale tutti costanti, un quinto delle vetrine curate a mano puntate su prodotti cancellati, e lo stesso prodotto presente in media cinque volte. Nessuna di queste condizioni era nota a chi gestiva il servizio.

Il guadagno di qualita' viene dai dati, ed e' la parte piu' redditizia del progetto. Indicizzare i campi gia' presenti chiude un quinto delle ricerche senza risposta e non costa nulla. Generare il vocabolario dei clienti con un modello economico, sotto verifica automatica, recupera migliaia di ricerche mensili. Normalizzare il documento porta le posizioni utili del pannello dal 54% al 96% e la coerenza dei risultati dall'85% al 92%, senza modificare una sola impostazione del motore.

Liberata dal costo per interrogazione, la ricerca torna a essere un problema di progettazione anziche' di budget.

MI

Massimo Ivaldi

A SEARCH COMPANY

CONCLUSIONE

Il costo scende.

La qualità resta.

La migrazione non è una semplice sostituzione del motore. È l'occasione per rendere portabile il merchandising, correggere la struttura dei dati e riportare la ricerca sotto il controllo dell'azienda.

01 / Misurare

Costruire il corpus di prova dai registri di utilizzo reali.

02 / Trasferire

Ricostruire regole e condizioni sulla base dei comportamenti osservati.

03 / Verificare

Confrontare fedeltà, latenza e copertura prima del rilascio.

04 / Migliorare

Intervenire sui dati con modifiche misurate e reversibili.

massimoivaldi.com · A search company

settembre 2026