MI

Massimo Ivaldi

A SEARCH COMPANY

WHITE PAPER / BENCHMARK TECNICO · WOOCOMMERCE

Benchmark della ricerca WooCommerce 20.000 prodotti

La ricerca come sistema di crescita.



settembre 2026 · edizione 1.0

LA RICERCA COME SISTEMA DI CRESCITA.

Il risultato dipende anche dalla configurazione.

Quattro configurazioni di ricerca sono state valutate con le stesse 26 interrogazioni, sullo stesso catalogo e sulla stessa infrastruttura. Il confronto separa l'effetto dello strumento da quello della modellazione dei dati.

20.000	167.584	26	94%
PRODOTTI DI ABBIGLIAMENTO	VARIANTI CON SKU	INTERROGAZIONI VALUTATE	RICHIAMO MEDIO MIGLIORE

DENTRO QUESTO BENCHMARK

01 ABSTRACT

02 OGGETTO E MOTIVAZIONE

03 PROTOCOLLO SPERIMENTALE

04 RISULTATI AGGREGATI

05 RISULTATI PER CASO

06 ANALISI PER CONFIGURAZIONE

07 ESPERIMENTO A VARIABILE SINGOLA

08 SCOMPOSIZIONE DEL DIFFERENZIALE

09 RICERCA PER SIMILARITÀ SEMANTICA

10 PROMOZIONE CONDIZIONATA DEI PRODOTTI

11 LIMITI DELLO STUDIO

12 CONCLUSIONI

13 APPENDICE A — RIPRODUCIBILITÀ

14 APPENDICE B — NOTAZIONE

La variabile indipendente

Le due configurazioni Meilisearch condividono motore, host, catalogo e interrogazioni. Il differenziale osservato dipende dalla modellazione dei documenti e dalla parametrizzazione dell'indice.

ABSTRACT

Abstract

È stato costruito un catalogo WooCommerce di 20.000 prodotti di abbigliamento con 167.584 varianti e sottoposto a un protocollo di valutazione costituito da 26 interrogazioni, applicato a quattro configurazioni di ricerca sulla medesima infrastruttura e nella medesima finestra temporale.

```
WooCommerce  Relevanssi   Meilisearch    Meilisearch
                      nativo      4.28.2    via CelerSearch   configurato
  esiti ok               2           5            10              18
  esiti parziali         4          19            12               8
  insiemi vuoti         20           2             4               0
  richiamo medio       19%         52%           66%             94%
  precisione@100       21%         62%           64%             78%
  latenza mediana    400 ms     7.758 ms        87 ms           87 ms
  latenza massima    555 ms    33.643 ms       385 ms          385 ms
  indicizzazione       n/d      8 h 40 min     2,4 ore        3 minuti
  dimensione indice    n/d       512 MB        140,7 MB       162,5 MB
  collocazione         n/d    nel DB del sito  esterna        esterna
```

Le due configurazioni di destra condividono motore, host, catalogo e insieme di interrogazioni. L'unica variabile indipendente è la modellazione dei documenti e la parametrizzazione dell'indice. Il differenziale osservato — 28 punti percentuali di richiamo e quattro insiemi di risultati vuoti — è pertanto integralmente attribuibile a tale variabile.

Questo è il risultato principale del presente studio.

01 / BENCHMARK

OGGETTO E MOTIVAZIONE

La funzione di ricerca interna di un catalogo online costituisce il canale attraverso il quale l'utente esplicita la propria intenzione d'acquisto in forma non mediata. A differenza della navigazione per categorie, che vincola l'utente alla tassonomia predisposta dal venditore, la ricerca testuale registra la formulazione spontanea del bisogno.

La letteratura sull'usabilità dei sistemi di e-commerce documenta da tempo la correlazione fra utilizzo della ricerca interna e tasso di conversione. La conseguenza operativa è che un difetto della funzione di ricerca non si manifesta come errore visibile: si manifesta come assenza di ordini, indistinguibile dal disinteresse.

L'ipotesi sottoposta a verifica in questo studio è articolata in tre punti:

```
H1  la ricerca nativa di WooCommerce presenta un difetto di copertura
      strutturale, non riconducibile a parametrizzazione
  H2  i plugin che operano sul database del sito riducono il difetto di
      copertura introducendo un costo in latenza e occupazione
  H3  la sostituzione del motore, di per sé, non è condizione sufficiente: una
      quota rilevante del risultato dipende dalla modellazione dei dati
```

02 / BENCHMARK

PROTOCOLLO SPERIMENTALE

2.1

INFRASTRUTTURA

Due istanze di pari dimensionamento presso il medesimo fornitore:

```
HOST APPLICATIVO   1 vCPU, 1.963 MB RAM, Ubuntu 22.04.5, kernel 5.15
                     WordPress 7.1, WooCommerce 11.1.0, PHP 8.3.32
                     MySQL 8.0.40 (buffer pool 384 MB), Apache 2.4.52 prefork
```

```
HOST MOTORE        1 vCPU, 1.900 MB RAM, Debian 13, kernel 6.12.94
                     Meilisearch 1.53.2
```

```
INTERCONNESSIONE   RTT medio 71,6 ms (min 70,7 — max 74,6), perdita 0%
```

Le due istanze risiedono in regioni differenti. In una configurazione di produzione co-locata il RTT sarebbe dell'ordine di 1-2 ms. Le latenze attribuite a Meilisearch incorporano pertanto una penalità di 71 ms non rappresentativa dell'esercizio. Il dato non è stato corretto: la sottostima delle prestazioni del sistema proposto costituisce una scelta conservativa a garanzia della validità delle conclusioni.

Si segnala inoltre che la collazione MySQL adottata (`utf8mb4_0900_ai_ci`) è insensibile ai segni diacritici. Il caso di prova sull'accento è pertanto superato da tutte le configurazioni indipendentemente dal motore, e non è attribuibile al merito di alcuna di esse.

2.2

POPOLAZIONE

Catalogo di 20.000 prodotti generato con parametri deterministici, riproducibile per costruzione.

```
tassonomia          35 categorie su tre livelli (genere > macro > tipo)
  distribuzione       Uomo 8.024 — Donna 7.198 — Bambino 2.404 — Sport 2.374
  fornitori           30
  assi di variazione  taglia (17 modalità), colore (20), materiale (12),
                      vestibilità (4), stagione (3)
  varianti            167.584, ciascuna con identificativo proprio
  disponibilità       19.294 disponibili — 706 esauriti
  prezzo medio        133,58 EUR
```

Una proprietà della popolazione assume rilievo nell'analisi successiva: il prodotto cartesiano degli assi di variazione non è saturo. Su un campione di 300 prodotti variabili, a fronte di 8.237 combinazioni teoricamente ammissibili sono state generate 6.201 varianti effettive, per un tasso di saturazione del 75,3%. Il 21% dei prodotti

variabili presenta almeno una combinazione mancante.

La condizione riproduce il comportamento di un catalogo reale, nel quale la disponibilità per taglia è funzione del colore e viceversa.

2.3

INSIEME DELLE INTERROGAZIONI

Ventisei interrogazioni, definite prima di qualunque rilevazione e mantenute invariate per l'intera durata dello studio. Il congelamento preventivo dell'insieme è condizione necessaria: una revisione successiva all'osservazione dei risultati configurerebbe un adattamento del criterio al sistema misurato.

Le interrogazioni sono stratificate in tre famiglie secondo il grado di complessità sintattica e il numero di attributi coinvolti. L'elenco integrale, con i risultati per caso, è riportato alla sezione 4.

```
FAMIGLIA BASE (n=9)            interrogazioni a uno o due termini su attributi
                                 identificativi: codice prodotto, codice
                                 variante, denominazione fornitore, categoria,
                                 attributo singolo. Include due casi di
                                 robustezza lessicale (refuso, flessione).
```

```
FAMIGLIA CODA LUNGA (n=9)      interrogazioni da tre a cinque termini
                                 combinanti tipologia merceologica e attributi.
                                 Riproduce il comportamento dell'utente che
                                 conosce l'oggetto cercato.
```

```
FAMIGLIA CONVERSAZIONALE (n=8)  interrogazioni in forma di sintagma naturale,
                                 comprensive di preposizioni, articoli e
                                 riempitivi. Riproduce il comportamento indotto
                                 dalla diffusione delle interfacce assistite.
```

2.4

VARIABILI DIPENDENTI

Sono state adottate due metriche congiunte.

```
RICHIAMO          |R ∩ P| / |R|      quota dell'insieme di riferimento
                                       effettivamente restituita
  PRECISIONE@100  |R ∩ P₁₀₀| / 100   quota dei primi cento elementi restituiti
                                       appartenenti all'insieme di riferimento
```

dove R è l'insieme di riferimento e P l'insieme restituito dal sistema.

L'adozione congiunta è metodologicamente necessaria. Il richiamo, isolato, è massimizzabile in modo banale: un operatore disgiuntivo restituisce l'intera popolazione e consegue richiamo unitario. La precisione, isolata, è massimizzabile restituendo un solo elemento corretto. Solo la lettura congiunta consente di distinguere un

sistema che risponde da un sistema che si limita a produrre output.

La distinzione non è accademica. Una delle configurazioni misurate produce, su un'interrogazione la cui risposta corretta consta di 7 elementi, un insieme di 1.471 elementi con precisione@100 nulla. Una metrica di solo richiamo classificherebbe tale comportamento come parzialmente riuscito.

2.5

COSTRUZIONE DELL'INSIEME DI RIFERIMENTO

L'insieme di riferimento è determinato per ciascuna interrogazione mediante interrogazione strutturata diretta sul database relazionale, indipendente dal sistema in valutazione. I criteri combinano appartenenza tassonomica (`term_relationships`) e corrispondenza testuale su titolo.

La scelta comporta una limitazione nota, discussa alla sezione 10.

2.6

CONTROLLI DI VALIDITÀ INTERNA

Prima dell'accettazione di ciascuna serie di misure è stata eseguita un'interrogazione di controllo a risposta nota, al fine di verificare che il sistema interrogato fosse effettivamente quello in valutazione.

Il controllo non è pleonastico. Una delle configurazioni provate ignorava in modo silente le interrogazioni programmatiche, restituendo i risultati della ricerca nativa. La serie di misure risultante era internamente coerente e statisticamente plausibile, e sarebbe stata accettata in assenza del controllo.

---

03 / BENCHMARK

RISULTATI AGGREGATI

WooCommerce	Relevanssi	CelerSearch	Meilisearch cfg	
esiti ok	2	5	10	18
esiti parziali	4	19	12	8
insiemi vuoti	20	2	4	0
richiamo medio	19%	52%	66%	94%
precisione@100 media	21%	62%	64%	78%
latenza mediana	400 ms	7.758 ms	87 ms	87 ms
latenza massima	555 ms	33.643 ms	385 ms	385 ms

Criterio di classificazione degli esiti:

```
ok         richiamo ≥ 0,90 e precisione ≥ 0,50
  parziale   insieme non vuoto, criterio ok non soddisfatto
  vuoto      cardinalità nulla dell'insieme restituito
```

Le quattro serie sono state rilevate nella medesima giornata, con identica configurazione del database relazionale, sul medesimo catalogo.

---

04 / BENCHMARK

RISULTATI PER CASO

Si riporta l'insieme integrale delle interrogazioni con gli esiti delle quattro configurazioni. Notazione: n = cardinalità dell'insieme restituito; R = richiamo percentuale; P = precisione@100 percentuale. La colonna "rif." indica la cardinalità dell'insieme di riferimento.

FAMIGLIA BASE

```
WooComm.   Relevanssi
  interrogazione                rif.       n   R    P       n    R    P
```

```
POL-UOM-05000                     1       0   0    0       1  100  100
  CAP-UOM-00000-ECRU-0            1       0   0    0       0    0    0
  Atelier Della Brianza         680       0   0    0     500   74  100
  Abbigliamento               4.827       0   0    0     500    4   40
  Blu Navy Cashmere             216       0   0    0      83   31   80
  camica                        822       0   0    0       0    0    0
  camicia                     1.268     844  67  100     500   39  100
  ecru                          334     334 100  100     500  100  100
  lino                        2.163   1.666  77  100     500   23  100
```

```
CelerSearch  Meilisearch cfg
  interrogazione                rif.       n    R    P       n    R    P
```

```
POL-UOM-05000                     1       3  100   33       1  100  100
  CAP-UOM-00000-ECRU-0            1       2  100   50       1  100  100
  Atelier Della Brianza         680     680  100  100     680  100  100
  Abbigliamento               4.827   1.000    7   31   6.000   42   31
  Blu Navy Cashmere             216     269  100   96     269  100   96
  camica                        822   1.000   78   69   1.662  100   69
  camicia                     1.268     844   67  100     844   67  100
  ecru                          334   1.000  100  100   1.955  100  100
  lino                        2.163   1.000   34   70   2.803  100  100
```

FAMIGLIA CODA LUNGA

```
WooComm.   Relevanssi
  interrogazione                rif.       n   R    P       n    R    P
```

```
camicia oxford blu navy           3       0   0    0       1   33  100
  maglione cashmere donna          34       0   0    0      39   94   82
```

```
  pantalone chino cotone           10       11  60   55       5   20   40
  giacca trench nero                6        0   0    0       1   17  100
  pantalone cargo denim            12       10  83  100       7   58  100
  giacca bomber uomo               13        0   0    0       4   31  100
  abito lino estivo donna           2        0   0    0       3   50   33
  cappotto doppiopetto lana merino  1        1 100  100       1  100  100
  sciarpa cashmere grigio perla     6        0   0    0       2   33  100
```

```
CelerSearch  Meilisearch cfg
  interrogazione                rif.       n    R    P       n    R    P
```

```
camicia oxford blu navy           3       3  100  100       3  100  100
  maglione cashmere donna          34      42  100   81      42  100   81
  pantalone chino cotone           10     227  100   10     227  100   10
  giacca trench nero                6       6  100  100       6  100  100
  pantalone cargo denim            12      16  100   75      16  100   75
  giacca bomber uomo               13      13  100  100      13  100  100
  abito lino estivo donna           2       5  100   40       5  100   40
  cappotto doppiopetto lana merino  1       5  100   20       5  100   20
  sciarpa cashmere grigio perla     6       6  100  100       6  100  100
```

FAMIGLIA CONVERSAZIONALE

```
WooComm.   Relevanssi
  interrogazione                rif.       n   R    P       n    R    P
```

```
jeans uomo taglia 40             71       0   0    0   1.447   72   17
  vestito di seta rosso da donna    7       0   0    0   1.471   29    0
  camicia bianca di lino            9       0   0    0     983   67    1
  cappotto da uomo taglia 52       24       0   0    0   1.238  100   13
  maglione nero in lana merino      8       0   0    0       8  100  100
  gonna plissettata taglia 42       5       0   0    0     984   40    2
  qualcosa in cashmere per l...   708       0   0    0     173   24  100
  polo azzurra di cotone da uomo    5       0   0    0   1.449  100    1
```

```
CelerSearch  Meilisearch cfg
  interrogazione                rif.       n    R    P       n    R    P
```

```
jeans uomo taglia 40             71      17   24  100      71  100  100
  vestito di seta rosso da donna    7       0    0    0       7  100  100
  camicia bianca di lino            9       3   33  100      11  100   82
  cappotto da uomo taglia 52       24       0    0    0      76  100   32
  maglione nero in lana merino      8       3   38  100       8  100  100
  gonna plissettata taglia 42       5       2   40  100       6  100   83
  qualcosa in cashmere per l...   708       0    0    0     350   35   90
  polo azzurra di cotone da uomo    5       0    0    0      21  100   24
```

05 / BENCHMARK

ANALISI PER CONFIGURAZIONE

5.1

WOOCOMMERCE NATIVO

Venti interrogazioni su ventisei producono un insieme vuoto. Le sei che producono risultati sono, senza eccezione, quelle nelle quali il termine cercato compare letteralmente nel titolo del prodotto.

La ricerca nativa interroga i campi `post_title` e `post_excerpt`. Non accede agli identificativi commerciali, agli attributi di variazione né alle tassonomie. L'ipotesi H1 risulta pertanto verificata: il difetto è di dominio, non di parametrizzazione, e non esiste configurazione di WooCommerce che lo corregga.

Si rileva inoltre una latenza di 3,85 secondi per il rendering completo della pagina dei risultati, di cui la componente attribuibile all'interrogazione del database è frazione minoritaria.

5.2

RELEVANSSI 4.28.2

Relevanssi costruisce un indice invertito in una tabella dedicata del database del sito, includendo attributi, tassonomie e campi personalizzati. Il difetto di copertura si riduce da 20 a 2 insiemi vuoti; il richiamo medio passa dal 19% al 52%. L'ipotesi H2 risulta verificata nella sua prima parte.

I costi introdotti sono di tre ordini.

LATENZA. Mediana 7.758 ms, massimo 33.643 ms, 19 casi su 26 oltre i 5 secondi. L'architettura prevede due interrogazioni SQL per ciascun termine — una per l'insieme dei documenti, una per la frequenza documentale — con ricombinazione in spazio applicativo. Un sintagma di quattro termini comporta otto interrogazioni.

È stata verificata l'indipendenza della latenza dalle risorse disponibili: triplicando il buffer pool InnoDB da 128 a 384 MB la mediana si riduce del 16% (da 9.168 a 7.758 ms) e il massimo resta invariato. Il collo di bottiglia non è l'I/O.

OCCUPAZIONE. L'indicizzazione di 187.584 documenti richiede 8 ore e 40 minuti e produce 1.598.117 righe per 512 MB, collocati nel database applicativo. La tabella risultante è la seconda per dimensione dell'intero schema, superata solo da `wp_postmeta` e pari a cinque volte `wp_posts`. Gli indici della tabella (275 MB) eccedono i dati (237 MB), conseguenza della necessità di supportare anche la corrispondenza suffissale mediante una colonna di termini invertiti.

DEGRADO DELLA PRECISIONE. È il fenomeno di maggior rilievo. Relevanssi prevede un meccanismo di ripiego che, in assenza di risultati sotto congiunzione, ripete l'interrogazione in disgiunzione senza segnalazione all'applicazione chiamante. L'effetto sulla famiglia conversazionale è documentato nella tabella della sezione 4: quattro casi su otto presentano precisione@100 non superiore a 0,02 su insiemi di cardinalità compresa fra 983 e 1.471.

La rilevanza del fenomeno è metodologica prima che pratica: sotto una metrica di solo richiamo, questi quattro casi risulterebbero fra i migliori della serie.

LIMITI ULTERIORI. Sono stati rilevati per ispezione del codice e verificati sui dati:

```
troncamento      un parametro attivo per impostazione predefinita limita a 500
                   i documenti restituiti per termine, ordinati per frequenza
                   grezza prima dell'applicazione del punteggio di rilevanza. Per
                   un termine con frequenza documentale 16.163 la frazione
                   esaminata è del 3%
```

```
congiunzione     l'operatore AND non è espresso come predicato SQL ma applicato
                   in spazio applicativo sulle liste già troncate
```

```
morfologia       assenza di riduzione alla radice per la lingua italiana.
                   Verificato: giacca 16.470 / giacche 2.370; pantalone 17.368 /
                   pantaloni 2.373; camicia 9.221 / camicie 1.149 costituiscono
                   termini distinti e non correlati
```

```
soglia lessicale il parametro di lunghezza minima, posto a 2, esclude dalla
                   tokenizzazione le stringhe di un carattere. Verificato per
                   conteggio: xs 5.537 — xl 5.834 — xxl 5.801 — 3xl 5.690;
                   s, m, l assenti dall'indice. L'interrogazione "taglia M" è
                   pertanto insoddisfacibile per costruzione
```

```
varianti         i 167.584 documenti di variante costituiscono l'89%
                   dell'indice, circa sette delle 8h40 di elaborazione e 450 dei
                   512 MB. Il documento restituito è la variante, che una query
                   di catalogo filtrata su `post_type = product` esclude.
                   L'interrogazione su codice di variante produce insieme vuoto
```

MODELLO DI RILEVANZA. Relevanssi adotta TF-IDF con elevazione al quadrato del fattore di frequenza documentale inversa. I valori misurati sulla popolazione:

```
termine              df        IDF Relevanssi    IDF BM25
    composizione     19.500             5,12          2,26
    polo              4.559            13,82          3,72
    identificativo        1           131,08         11,74
    rapporto raro/comune               25,6×          5,2×
```

L'amplificazione conferisce ai termini a bassa frequenza documentale un peso relativo quintuplo rispetto alla formulazione BM25. Il comportamento è ottimale sugli identificativi e subottimale sui sintagmi multi-attributo, nei quali i termini dovrebbero concorrere in misura comparabile.

Si ritiene tuttavia necessario precisare che la sostituzione della funzione di ordinamento non sarebbe risolutiva. La normalizzazione per lunghezza documentale prevista da BM25 favorisce i documenti brevi; nella popolazione in esame i documenti brevi sono le varianti, con lunghezza media di 4,9 termini contro i 39,1 dei prodotti, e cardinalità otto volte superiore. A parità di corrispondenza, BM25 collocherebbe sistematicamente le varianti al di sopra dei prodotti.

La conclusione che se ne trae è di ordine più generale: la criticità non risiede nella funzione di ordinamento ma nel modello dei dati. Un catalogo non è un corpus testuale, e gli attributi di variazione non sono termini: sono variabili categoriche a dominio finito.

5.3

MEILISEARCH 1.53.2

Meilisearch è un motore di ricerca open source implementato in Rust, progettato per l'esecuzione come servizio indipendente dall'applicazione. La differenza architetturale rispetto a un plugin che opera sul database applicativo consiste nella separazione della struttura di ricerca dalla struttura di persistenza: un indice invertito ottimizzato per l'accesso per termine, in luogo di un insieme di tabelle relazionali interrogate per corrispondenza testuale.

La latenza mediana misurata è di 87 ms, dei quali 71 imputabili all'interconnessione di rete. Il tempo di elaborazione dichiarato dal motore è compreso fra 1 e 8 ms.

Le funzionalità rilevanti ai fini della valutazione comprendono: tolleranza all'errore di battitura parametrizzabile per campo; predicati di filtro su attributi strutturati; distribuzione delle faccette calcolata contestualmente all'interrogazione; criteri di ordinamento multipli; dizionari di sinonimi e di termini non significativi; regole di promozione condizionate all'interrogazione; ricerca per similarità semantica con inferenza locale.

5.4

CELERSEARCH 1.7.2

CelerSearch è un plugin WordPress che instrada la ricerca verso un'istanza Meilisearch esterna. Non include il motore.

La configurazione con parametri predefiniti consegue 10 esiti ok, 12 parziali e 4 insiemi vuoti, con richiamo medio 66% e precisione@100 64%. Il confronto con Relevanssi è nettamente favorevole su tutti gli indicatori: dimezzamento degli insiemi vuoti, incremento di 14 punti del richiamo, riduzione della latenza mediana di un fattore 89.

L'interfaccia di configurazione espone l'insieme completo dei parametri dell'indice: regole di ordinamento, attributi ricercabili, filtrabili e ordinabili, sinonimi, termini non significativi, tolleranza all'errore, prossimità, frammentazione dei documenti, limiti di paginazione, regole di promozione, ricerca semantica.

Sono state rilevate due criticità di natura non parametrica.

EFFICIENZA DI INDICIZZAZIONE. La trasmissione dei documenti avviene con cardinalità unitaria per richiesta HTTP (verificato: `receivedDocuments = 1` su tutte le operazioni osservate). Meilisearch è ottimizzato per l'ingestione in lotti. Il differenziale misurato a parità di motore e popolazione:

```
lotti da 2.000 documenti      24 secondi per 20.000 documenti
    cardinalità unitaria          2,4 ore per 20.000 documenti
```

Fattore 354. Il parametro di dimensione del lotto esposto dal plugin governa la cardinalità dei post caricati per ciclo applicativo, non quella dei documenti per richiesta. La scelta è motivata nel codice dalla volontà di mantenere un unico percorso di scrittura per tutte le operazioni — sincronizzazione in tempo reale, coda

differita, ricostruzione completa — ciascuna delle quali appone una marca temporale utilizzata per la successiva rimozione dei documenti orfani. La motivazione è tecnicamente fondata; il costo su popolazioni di questa cardinalità è elevato.

Si rileva inoltre che il ciclo di indicizzazione è pilotato dal browser dell'operatore mediante chiamate asincrone: l'interruzione della sessione interrompe l'elaborazione.

POPOLAMENTO DEL CAMPO VARIANTI. Il documento prodotto prevede un campo destinato agli identificativi di variante. A indicizzazione completata tale campo risulta vuoto su tutti i documenti esaminati, pur in presenza di un contatore di varianti valorizzato. Verifica su tre identificativi di variante del medesimo prodotto: due producono insieme vuoto; il terzo produce un risultato per coincidenza lessicale, essendo la denominazione del colore presente nel titolo del prodotto.

---

06 / BENCHMARK

ESPERIMENTO A VARIABILE SINGOLA

Il confronto fra le due configurazioni Meilisearch costituisce il nucleo dello studio, poiché consente di isolare il contributo della modellazione dei dati dal contributo dello strumento.

```
CONTROLLATE     motore (Meilisearch 1.53.2), host, popolazione (20.000
                  prodotti), insieme delle interrogazioni, insieme di
                  riferimento, finestra temporale, condizioni di rete
```

```
MANIPOLATA      struttura del documento indicizzato e parametrizzazione
                  dell'indice
```

```
CelerSearch      configurazione analitica
    insiemi vuoti              4                   0
    richiamo medio            66%                 94%
    precisione@100            64%                 78%
    latenza mediana          87 ms               87 ms
    indicizzazione          2,4 ore             3 minuti
```

L'invarianza della latenza conferma l'identità del motore e la validità del controllo. Il differenziale su richiamo e precisione è pertanto integralmente attribuibile alla variabile manipolata. L'ipotesi H3 risulta verificata.

Si procede alla scomposizione del differenziale nelle sue componenti.

---

07 / BENCHMARK

SCOMPOSIZIONE DEL DIFFERENZIALE

7.1

DIZIONARIO DEI TERMINI NON SIGNIFICATIVI

I quattro insiemi vuoti della configurazione predefinita appartengono tutti alla famiglia conversazionale:

```
vestito di seta rosso da donna
    cappotto da uomo taglia 52
    qualcosa in cashmere per l'inverno
    polo azzurra di cotone da uomo
```

In assenza di dizionario dei termini non significativi e sotto strategia di corrispondenza congiuntiva, le preposizioni e i riempitivi entrano nell'insieme dei termini obbligatori. La probabilità che un documento di catalogo contenga la preposizione "da" nel proprio testo indicizzato è prossima a zero; l'insieme risultante è vuoto per costruzione.

Il dizionario adottato nella configurazione analitica comprende 60 forme italiane. Il parametro è esposto dall'interfaccia di CelerSearch e vuoto per impostazione predefinita.

7.2

DIZIONARIO DEI SINONIMI

La popolazione non contiene alcuna occorrenza del lemma "vestito": la denominazione adottata è "abito". Un'interrogazione contenente il primo termine non è soddisfacibile in assenza di una relazione di equivalenza dichiarata.

Effetto misurato sull'interrogazione "vestito di seta rosso da donna", con 13 relazioni di equivalenza dichiarate: da insieme vuoto a 7 elementi su 7 attesi, precisione@100 unitaria.

7.3

ESPOSIZIONE TESTUALE DEGLI ATTRIBUTI

L'interrogazione "taglia 40" contiene un termine — l'etichetta dell'attributo — che non compare in alcun documento. Il valore è presente come modalità di una variabile categorica; l'etichetta appartiene allo schema, non ai dati.

La configurazione analitica genera in fase di indicizzazione un campo testuale che accoppia etichetta e valore:

```
colore Bordeaux · colore Verde Oliva · taglia 40 · taglia 42 ·
    materiale Seta · vestibilità Slim · stagione Continuativo
```

Costo di archiviazione: alcune centinaia di byte per documento. Effetto misurato sull'interrogazione "jeans uomo taglia 40": da 17 elementi su 71 attesi (richiamo 24%) a 71 su 71 (richiamo 100%), con precisione@100 unitaria in entrambi i casi.

7.4

AGGREGAZIONE DEGLI IDENTIFICATIVI DI VARIANTE

L'identificativo riportato sull'etichetta fisica del capo è quello della variante, non del prodotto. Costituisce la chiave di ricerca utilizzata dall'acquirente in possesso dell'articolo, dall'operatore di magazzino e dal rivenditore.

Le configurazioni che indicizzano la variante come documento autonomo producono insiemi vuoti su query di catalogo, che filtrano per tipo di post. Le configurazioni che non la indicizzano affatto producono lo stesso esito.

La configurazione analitica aggrega i 167.584 identificativi di variante nei 20.000 documenti di prodotto che li contengono. L'effetto secondario è rilevante: la cardinalità dell'indice si riduce da 187.584 a 20.000 documenti, e il tempo di costruzione da 8 ore e 40 minuti a 3 minuti.

7.5

RAPPRESENTAZIONE DELLE COMBINAZIONI ESISTENTI

È la componente di maggior interesse analitico e la meno frequentemente rilevata.

L'applicazione di due predicati di filtro su attributi multivalore — colore e taglia — produce, in assenza di una rappresentazione della combinazione, un insieme che comprende i documenti che soddisfano i due predicati su elementi distinti dei rispettivi array.

Formalmente: dato un prodotto p con insiemi di colori C(p) e taglie T(p), e dato l'insieme V(p) delle varianti effettivamente esistenti, il predicato `c ∈ C(p) ∧ t ∈ T(p)` è condizione necessaria ma non sufficiente per `(c,t) ∈ π(V(p))`.

La differenza è misurabile:

```
predicato              filtri indipendenti   combinazione   differenza
    Bordeaux ∧ 40                  257              249            8
    Blu Navy ∧ 42                  245              229           16
    Nero ∧ M                       367              353           14
    Verde Oliva ∧ 44               254              237           17
    Rosso ∧ L                      353              334           19
```

```
totale                      1.476            1.402      74 (5,0%)
```

Il 5,0% dei risultati restituiti da un filtro combinato non corrisponde a un'articolazione disponibile. L'effetto sull'esperienza d'acquisto è la selezione di una combinazione esaurita.

La correzione richiede la generazione, in fase di indicizzazione, di una chiave composta per ciascuna variante esistente, e il suo impiego come predicato di filtro in luogo della congiunzione dei predicati singoli. Non è un parametro configurabile: è una decisione di modellazione che presuppone l'analisi della distribuzione delle varianti nella popolazione.

7.6

ORDINAMENTO DEI CAMPI RICERCABILI

Il caso è di piccola entità e alto valore illustrativo.

L'interrogazione "lino" produce, nelle due configurazioni, insiemi di identica cardinalità (2.803 elementi). La precisione@100 differisce di 30 punti: 70% contro 100%.

La causa risiede nell'ordine dichiarato degli attributi ricercabili, che governa la regola di ordinamento per attributo. Nella configurazione predefinita la denominazione del fornitore risiede nel medesimo contenitore degli attributi di prodotto e concorre con peso equivalente; i prodotti del fornitore denominato "Cotone e Lino Sud" competono in classifica con quelli effettivamente realizzati in lino. Nella configurazione analitica il fornitore costituisce campo distinto, collocato successivamente agli attributi nell'ordine di priorità.

---

08 / BENCHMARK

RICERCA PER SIMILARITÀ SEMANTICA

Esiste una classe di interrogazioni per le quali nessuna configurazione lessicale può produrre risultati, poiché i termini impiegati non appartengono al vocabolario del catalogo:

```
qualcosa di caldo per l'inverno           insieme vuoto
    cosa metto a un matrimonio                insieme vuoto
    un regalo elegante per lei                insieme vuoto
    vestiti per andare a correre              insieme vuoto
    capo traspirante per la palestra          insieme vuoto
```

L'insieme di riferimento non è vuoto: la popolazione contiene capospalla in lana merino, abiti da cerimonia e articoli tecnici. I termini "caldo", "matrimonio", "correre" non compaiono tuttavia in alcun documento.

La ricerca per similarità semantica opera su rappresentazioni vettoriali dense. Ciascun documento e ciascuna interrogazione sono proiettati in uno spazio metrico; il criterio di ordinamento è la prossimità nello spazio. La relazione fra "matrimonio" e "abito da cerimonia" è codificata nella geometria dello spazio, non nella sovrapposizione lessicale.

Meilisearch incorpora il motore di inferenza nel proprio eseguibile e supporta l'esecuzione locale del modello, senza dipendenza da servizi esterni, chiavi di accesso o costi per interrogazione. Le alternative supportate comprendono provider esterni, server di inferenza propri e vettori calcolati a monte.

MISURE. Generazione delle rappresentazioni per 20.000 documenti: 85 minuti su un singolo core. Incremento dell'indice: da 162,5 a 263,4 MB. Latenza di interrogazione in modalità ibrida: 42-51 ms.

Le cinque interrogazioni producono risultati pertinenti: "cosa metto a un matrimonio" restituisce documenti della categoria Abiti; "qualcosa di caldo per l'inverno" restituisce articoli in pile e a denominazione termica; "vestiti per andare a correre" restituisce leggings.

CRITICITÀ DI CONFIGURAZIONE RILEVATE. Quattro, nessuna delle quali documentata in posizione evidente:

```
a) il modello predefinito è addestrato su corpus anglofono. L'impiego su
     catalogo italiano produce rappresentazioni non discriminanti, senza alcuna
     segnalazione
```

```
b) i modelli non predefiniti richiedono la specificazione esplicita
     dell'identificativo di revisione; in sua assenza l'operazione fallisce con
     errore di recupero risorsa, non diagnostico della causa
```

```
c) la classe di addestramento del modello è determinante. I modelli addestrati
     su similarità fra frasi (paraphrase) producono risultati non pertinenti dove
     sono richiesti modelli addestrati su recupero asimmetrico domanda-documento
```

```
d) gli accorgimenti di formulazione documentati per taluni modelli, applicati
     conformemente alla specifica, degradano i risultati e compromettono la
     componente lessicale della ricerca ibrida
```

---

09 / BENCHMARK

PROMOZIONE CONDIZIONATA DEI PRODOTTI

Un motore esterno consente la definizione di regole di promozione condizionate all'interrogazione: l'assegnazione di una posizione fissa a documenti selezionati in presenza di determinati termini.

Misura. Sull'interrogazione "cappotto", l'ordinamento per sola rilevanza testuale collocava in prima posizione un articolo con 26 unità vendute. L'applicazione di una regola di promozione sul prodotto a maggior volume colloca in prima posizione l'articolo con 1.358 unità.

AVVERTENZA METODOLOGICA. I dati di vendita, valutazione e sconto non erano presenti nella popolazione generata e sono stati simulati con distribuzione a coda lunga, deterministica e riproducibile, al solo fine di rendere dimostrabile la funzione. I valori citati in questa sezione sono verosimili ma non reali.

Si segnala che la funzione, nell'implementazione offerta dal plugin gratuito, non era operativa sulla versione corrente del motore: il pannello consentiva la composizione della regola ma l'operazione di salvataggio falliva per quattro incompatibilità di schema fra la specifica implementata e quella corrente (denominazione di un campo, struttura del contenitore delle condizioni, forma di una condizione, base dell'indicizzazione posizionale). La composizione manuale della regola nella forma corrente produce il comportamento atteso.

L'osservazione ha rilievo generale: la dipendenza da un componente di terze parti per una funzione a impatto commerciale diretto comporta un rischio di disallineamento fra i cicli di rilascio del motore e quelli del componente.

10 / BENCHMARK

LIMITI DELLO STUDIO

Si esplicitano le limitazioni note, in quanto incidono sulla generalizzabilità delle conclusioni e operano in direzione sfavorevole alle tesi sostenute.

10.1

NATURA SINTETICA DELLA POPOLAZIONE

I prodotti sono verosimili nella struttura — tassonomia, attributi, varianti, identificativi — ma non nel contenuto editoriale. Le descrizioni sono generate da un vocabolario chiuso: su 187.920 termini distinti presenti nell'indice Relevanssi, 187.584 compaiono in un solo documento (gli identificativi), residuando un vocabolario condiviso di 336 forme.

La limitazione incide in misura determinante sulla sezione 8. Le rappresentazioni vettoriali sono computate su testo largamente ripetitivo, con conseguente compressione dello spazio metrico: il differenziale di punteggio fra la prima e la cinquemillesima posizione è risultato pari a 0,013. Le misure di ricerca semantica vanno pertanto interpretate come dimostrazione di funzionamento del meccanismo e non come stima della qualità conseguibile su popolazione reale.

Si rileva inoltre che la generazione combinatoria produce associazioni merceologicamente incoerenti: il 12% dei prodotti presenta materiali non compatibili con la tipologia (177 cinture in cashmere, 165 in denim, 161 in pile). Un sistema di ricerca semantica riflette fedelmente tali associazioni.

10.2

SOTTOSTIMA DELLA PRECISIONE

L'insieme di riferimento è definito per corrispondenza testuale su titolo congiunta ad appartenenza tassonomica. Il criterio è stato formulato quando l'unico sistema in valutazione operava per corrispondenza su titolo, e penalizza i sistemi che accedono alla struttura del catalogo.

Caso documentato: l'interrogazione "pantalone chino cotone" produce 227 elementi classificati con precisione@100 pari a 0,10. L'esame dell'insieme restituito mostra 205 articoli di tipologia Chino, 17 Pantalone e 5 Chinos, tutti appartenenti alla categoria Pantaloni e con materiale Cotone o Cotone Organico. I risultati sono pertinenti; l'insieme di riferimento richiedeva la compresenza di entrambi i lemmi nel titolo ed escludeva la modalità "Cotone Organico".

Il criterio non è stato modificato successivamente all'osservazione dei risultati. La revisione di un criterio di valutazione dopo l'osservazione degli esiti costituisce adattamento del criterio al sistema misurato.

10.3

SIMULAZIONE DEI SEGNALI COMMERCIALI

Come indicato alla sezione 9.

10.4

ASSENZA DI PROVE DI CARICO

Le rilevazioni sono state condotte in condizione di utente singolo. Il comportamento sotto concorrenza, su host a singolo core, non è stato valutato.

10.5

SINGOLA REPLICAZIONE

Non è stata stimata la varianza fra esecuzioni ripetute su infrastrutture distinte. Le latenze riportate sono mediane su singola serie.

10.6

DIFETTO NOTO NELLA CONFIGURAZIONE ANALITICA

L'interrogazione "taglia M" produce 16.189 elementi a fronte di 4.258 attesi. La causa è la corrispondenza prefissale applicata all'ultimo termine dell'interrogazione: il carattere "M" isolato corrisponde a qualunque forma che lo abbia come prefisso, condizione soddisfatta da tutti i documenti della popolazione. L'ordinamento mitiga l'effetto — i primi venti elementi presentano tutti la modalità M — ma la cardinalità dichiarata è errata. La correzione richiede il riconoscimento dei valori di taglia in fase di interpretazione dell'interrogazione e la loro traduzione in predicato di filtro.

---

11 / BENCHMARK

CONCLUSIONI

11.1 L'ipotesi H1 è verificata. La ricerca nativa di WooCommerce presenta un

> difetto di dominio: non accede agli attributi di variazione, alle tassonomie né agli identificativi commerciali. Su una popolazione con varianti, ciò comporta l'insoddisfacibilità della maggioranza delle interrogazioni commercialmente rilevanti — venti su ventisei nella presente rilevazione. Il difetto non è correggibile per parametrizzazione.

11.2 L'ipotesi H2 è verificata. I plugin operanti sul database applicativo

> riducono il difetto di copertura (da 20 a 2 insiemi vuoti) introducendo costi in latenza (mediana 7.758 ms) e occupazione (512 MB nel database del sito). Si rileva inoltre un meccanismo di ripiego non segnalato che produce insiemi di cardinalità elevata e precisione prossima a zero, classificabili come successo da metriche di solo richiamo.

11.3 La sostituzione del motore risolve simultaneamente latenza e occupazione,

```
con differenziali di ordine di grandezza: 87 ms contro 7.758, indice
      esterno al database applicativo, costruzione in 3 minuti contro 8 ore e 40.
```

11.4 L'ipotesi H3 è verificata, ed è il risultato principale. A parità di

```
motore, host, popolazione e interrogazioni, il differenziale fra
      configurazione predefinita e configurazione analitica è di 28 punti di
      richiamo e 4 insiemi vuoti. Il differenziale è integralmente attribuibile
      a sei decisioni di modellazione e parametrizzazione, documentate alla
      sezione 7 e singolarmente quantificate.
```

11.5 Nessuna delle sei decisioni è determinabile in modo automatico, poiché

```
ciascuna dipende da proprietà specifiche della popolazione: lingua,
      convenzioni di denominazione adottate, assi di variazione impiegati,
      lessico del settore merceologico, distribuzione delle varianti.
```

Se ne deriva che l'adozione di un motore di ricerca esterno costituisce condizione necessaria ma non sufficiente. La componente residua — quantificata in 28 punti di richiamo, corrispondenti alla classe di interrogazioni a maggiore intenzione d'acquisto — richiede un'analisi preliminare della popolazione e una progettazione esplicita dello schema documentale.

APPENDICE

APPENDICE A — RIPRODUCIBILITÀ

La popolazione è generata con parametri deterministici (seme 20260906) ed è ricostruibile in forma identica. L'insieme delle 26 interrogazioni, i criteri di costruzione dell'insieme di riferimento e le due metriche sono documentati alle sezioni 2.3-2.5 e applicabili a qualunque catalogo WooCommerce.

Versioni impiegate: WordPress 7.1, WooCommerce 11.1.0, PHP 8.3.32, MySQL 8.0.40, Relevanssi 4.28.2, CelerSearch 1.7.2, Meilisearch 1.53.2.

APPENDICE

APPENDICE B — NOTAZIONE

```
R              insieme di riferimento per una data interrogazione
  P              insieme restituito dal sistema
  P₁₀₀           primi cento elementi di P secondo l'ordinamento del sistema
  richiamo       |R ∩ P| / |R|
  precisione@100 |R ∩ P₁₀₀| / min(100, |P|)
  df             frequenza documentale di un termine
  n              cardinalità dell'insieme restituito
```

MI

Massimo Ivaldi

A SEARCH COMPANY

CONCLUSIONE

La ricerca non è una casella.
È un sistema.

Il motore conta. Ma il risultato nasce prima: dalla struttura del catalogo, dalle decisioni sui campi, dai sinonimi, dalle varianti e dal modo in cui l'intento reale delle persone viene tradotto in dati.

01 / Misurare

Partire da query reali, copertura e precisione. Non da impressioni.

02 / Modellare

Portare attributi, varianti e combinazioni nel posto giusto.

03 / Configurare

Trattare sinonimi, tolleranza e ranking come decisioni di business.

04 / Verificare

Rendere ogni scelta riproducibile, documentata e migliorabile.

massimoivaldi.com · A search company

settembre 2026