MI

Massimo Ivaldi

A SEARCH COMPANY

LIVRE BLANC / BENCHMARK TECHNIQUE · WOOCOMMERCE

Benchmark de la recherche WooCommerce 20 000 produits

La recherche comme système de croissance.



septembre 2026 · édition 1.0

LA RECHERCHE COMME SYSTÈME DE CROISSANCE.

Le résultat dépend aussi de la configuration.

Quatre configurations de recherche ont été évaluées à l'aide des mêmes 26 requêtes, sur le même catalogue et la même infrastructure. La comparaison distingue l'effet de l'outil de celui de la modélisation des données.

20 000	167 584	26	94%
PRODUITS D'HABILLEMENT	DÉCLINAISONS AVEC RÉFÉRENCE	REQUÊTES ÉVALUÉES	MEILLEUR RAPPEL MOYEN

DANS CE BENCHMARK

01 RÉSUMÉ

02 OBJET ET MOTIVATION

03 PROTOCOLE EXPÉRIMENTAL

04 RÉSULTATS AGRÉGÉS

05 RÉSULTATS PAR CAS

06 ANALYSE PAR CONFIGURATION

07 EXPÉRIENCE À VARIABLE UNIQUE

08 DÉCOMPOSITION DE L'ÉCART

09 RECHERCHE PAR SIMILARITÉ SÉMANTIQUE

10 MISE EN AVANT CONDITIONNELLE DES PRODUITS

11 LIMITES DE L'ÉTUDE

12 CONCLUSIONS

13 ANNEXE A — REPRODUCTIBILITÉ

14 ANNEXE B — NOTATION

La variable indépendante

Les deux configurations Meilisearch partagent le moteur, l'hôte, le catalogue et les requêtes. L'écart observé dépend de la modélisation des documents et du paramétrage de l'index.

RÉSUMÉ

Résumé

Un catalogue WooCommerce de 20 000 produits d'habillement comportant 167 584 déclinaisons a été constitué, puis soumis à un protocole d'évaluation de 26 requêtes appliqué à quatre configurations de recherche sur la même infrastructure et dans la même fenêtre temporelle.

```
WooCommerce  Relevanssi   Meilisearch    Meilisearch
                        natif      4.28.2    via CelerSearch   configuré
  résultats ok            2           5            10              18
  résultats partiels      4          19            12               8
  ensembles vides        20           2             4               0
  rappel moyen          19 %        52 %          66 %            94 %
  précision@100         21 %        62 %          64 %            78 %
  latence médiane      400 ms    7 758 ms        87 ms           87 ms
  latence maximale     555 ms   33 643 ms       385 ms          385 ms
  indexation            n/d      8 h 40 min      2,4 h         3 minutes
  taille de l'index     n/d       512 Mo        140,7 Mo       162,5 Mo
  emplacement           n/d    dans la BDD      externe         externe
```

Les deux configurations de droite partagent le moteur, l'hôte, le catalogue et l'ensemble des requêtes. La seule variable indépendante est la modélisation des documents et le paramétrage de l'index. L'écart observé — 28 points de pourcentage de rappel et quatre ensembles de résultats vides — est donc intégralement imputable à cette variable.

C'est le résultat principal de la présente étude.

01 / BENCHMARK

OBJET ET MOTIVATION

La fonction de recherche interne d'un catalogue en ligne constitue le canal par lequel l'utilisateur explicite son intention d'achat sans médiation. À la différence de la navigation par catégories, qui contraint l'utilisateur à la taxonomie prévue par le vendeur, la recherche textuelle enregistre la formulation spontanée du besoin.

La littérature relative à l'ergonomie des systèmes de commerce électronique documente de longue date la corrélation entre usage de la recherche interne et taux de conversion. La conséquence opérationnelle est qu'un défaut de la fonction de recherche ne se manifeste pas comme une erreur visible : il se manifeste comme une absence de commandes, indiscernable du désintérêt.

L'hypothèse soumise à vérification s'articule en trois propositions :

```
H1  la recherche native de WooCommerce présente un défaut de couverture
      structurel, non imputable au paramétrage
  H2  les extensions opérant sur la base de données du site réduisent ce défaut
      de couverture en introduisant un coût en latence et en occupation
  H3  la substitution du moteur n'est pas, en soi, une condition suffisante :
      une part significative du résultat dépend de la modélisation des données
```

02 / BENCHMARK

PROTOCOLE EXPÉRIMENTAL

2.1

INFRASTRUCTURE

Deux instances de dimensionnement équivalent chez le même fournisseur :

```
HÔTE APPLICATIF    1 vCPU, 1 963 Mo RAM, Ubuntu 22.04.5, noyau 5.15
                     WordPress 7.1, WooCommerce 11.1.0, PHP 8.3.32
                     MySQL 8.0.40 (buffer pool 384 Mo), Apache 2.4.52 prefork
```

```
HÔTE MOTEUR        1 vCPU, 1 900 Mo RAM, Debian 13, noyau 6.12.94
                     Meilisearch 1.53.2
```

```
INTERCONNEXION     RTT moyen 71,6 ms (min 70,7 — max 74,6), perte 0 %
```

Les deux instances résident dans des régions distinctes. Dans une configuration de production co-localisée, le RTT serait de l'ordre de 1 à 2 ms. Les latences imputées à Meilisearch intègrent donc une pénalité de 71 ms non représentative de l'exploitation. Cette donnée n'a pas été corrigée : la sous-estimation des performances du système proposé constitue un choix conservateur garantissant la validité des conclusions.

Il est en outre signalé que la collation MySQL retenue (`utf8mb4_0900_ai_ci`) est insensible aux signes diacritiques. Le cas d'essai portant sur l'accent est donc satisfait par toutes les configurations indépendamment du moteur, et n'est imputable au mérite d'aucune d'entre elles.

2.2

POPULATION

Catalogue de 20 000 produits généré avec des paramètres déterministes, reproductible par construction.

```
taxonomie          35 catégories sur trois niveaux (genre > macro > type)
  répartition        Homme 8 024 — Femme 7 198 — Enfant 2 404 — Sport 2 374
  fournisseurs       30
  axes de variation  taille (17 modalités), couleur (20), matière (12),
                     coupe (4), saison (3)
  déclinaisons       167 584, chacune dotée d'une référence propre
  disponibilité      19 294 disponibles — 706 épuisés
  prix moyen         133,58 EUR
```

Une propriété de la population prend une importance déterminante dans l'analyse ultérieure : le produit cartésien des axes de variation n'est pas saturé. Sur un échantillon de 300 produits déclinables, pour 8 237 combinaisons théoriquement admissibles, 6 201 déclinaisons effectives ont été générées, soit un taux de

saturation de 75,3 %. 21 % des produits déclinables présentent au moins une combinaison manquante.

Cette condition reproduit le comportement d'un catalogue réel, dans lequel la disponibilité par taille est fonction de la couleur et réciproquement.

2.3

ENSEMBLE DES REQUÊTES

Vingt-six requêtes, définies avant tout relevé et maintenues invariantes pendant toute la durée de l'étude. Le gel préalable de l'ensemble est une condition nécessaire : une révision postérieure à l'observation des résultats constituerait un ajustement du critère au système mesuré.

Les requêtes sont stratifiées en trois familles selon le degré de complexité syntaxique et le nombre d'attributs mobilisés. La liste intégrale, assortie des résultats par cas, figure en section 4.

Les chaînes interrogées sont conservées en italien : elles constituent les données primaires de l'expérience, le catalogue étant rédigé dans cette langue.

```
FAMILLE BASE (n=9)              requêtes d'un à deux termes portant sur des
                                  attributs identifiants : référence produit,
                                  référence de déclinaison, dénomination du
                                  fournisseur, catégorie, attribut isolé. Inclut
                                  deux cas de robustesse lexicale (faute de
                                  frappe, flexion).
```

```
FAMILLE TRAÎNE LONGUE (n=9)     requêtes de trois à cinq termes combinant
                                  typologie marchande et attributs. Reproduit le
                                  comportement de l'utilisateur qui connaît
                                  l'objet recherché.
```

```
FAMILLE CONVERSATIONNELLE (n=8)  requêtes sous forme de syntagme naturel,
                                   comportant prépositions, articles et termes de
                                   remplissage. Reproduit le comportement induit
                                   par la diffusion des interfaces assistées.
```

2.4

VARIABLES DÉPENDANTES

Deux métriques conjointes ont été retenues.

```
RAPPEL          |R ∩ P| / |R|     part de l'ensemble de référence
                                    effectivement restituée
  PRÉCISION@100   |R ∩ P₁₀₀| / 100  part des cent premiers éléments restitués
                                    appartenant à l'ensemble de référence
```

où R désigne l'ensemble de référence et P l'ensemble restitué par le système.

L'emploi conjoint est méthodologiquement nécessaire. Le rappel, pris isolément, est maximisable trivialement : un opérateur disjonctif restitue la population entière et atteint un rappel unitaire. La précision, prise isolément, est maximisable en ne restituant qu'un seul élément correct. Seule la lecture conjointe permet de distinguer un système qui répond d'un système qui se borne à produire une sortie.

La distinction n'est pas académique. L'une des configurations mesurées produit, sur une requête dont la réponse correcte comporte 7 éléments, un ensemble de 1 471 éléments avec une précision@100 nulle. Une métrique de rappel seul classerait ce comportement comme partiellement réussi.

2.5

CONSTRUCTION DE L'ENSEMBLE DE RÉFÉRENCE

L'ensemble de référence est déterminé pour chaque requête par interrogation structurée directe de la base de données relationnelle, indépendamment du système évalué. Les critères combinent appartenance taxonomique (`term_relationships`) et correspondance textuelle sur le titre.

Ce choix comporte une limitation connue, discutée en section 10.

2.6

CONTRÔLES DE VALIDITÉ INTERNE

Préalablement à l'acceptation de chaque série de mesures, une requête de contrôle à réponse connue a été exécutée afin de vérifier que le système interrogé était bien celui soumis à évaluation.

Ce contrôle n'est pas superflu. L'une des configurations essayées ignorait silencieusement les requêtes programmatiques et restituait les résultats de la recherche native. La série de mesures qui en résultait était cohérente en interne et statistiquement plausible, et aurait été acceptée en l'absence du contrôle.

---

03 / BENCHMARK

RÉSULTATS AGRÉGÉS

WooCommerce	Relevanssi	CelerSearch	Meilisearch cfg	
résultats ok	2	5	10	18
résultats partiels	4	19	12	8
ensembles vides	20	2	4	0
rappel moyen	19 %	52 %	66 %	94 %
précision@100 moyenne	21 %	62 %	64 %	78 %
latence médiane	400 ms	7 758 ms	87 ms	87 ms
latence maximale	555 ms	33 643 ms	385 ms	385 ms

Critère de classification des résultats :

```
ok        rappel ≥ 0,90 et précision ≥ 0,50
  partiel   ensemble non vide, critère ok non satisfait
```

```
  vide      cardinalité nulle de l'ensemble restitué
```

Les quatre séries ont été relevées le même jour, avec une configuration identique de la base de données relationnelle, sur le même catalogue.

---

04 / BENCHMARK

RÉSULTATS PAR CAS

L'ensemble intégral des requêtes est reporté avec les résultats des quatre configurations. Notation : n = cardinalité de l'ensemble restitué ; R = rappel en pourcentage ; P = précision@100 en pourcentage. La colonne « réf. » indique la cardinalité de l'ensemble de référence.

FAMILLE BASE

```
WooComm.   Relevanssi
  requête                       réf.       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
  requête                       réf.       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
```

FAMILLE TRAÎNE LONGUE

```
WooComm.   Relevanssi
  requête                       réf.       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
  requête                        réf.        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
```

FAMILLE CONVERSATIONNELLE

```
WooComm.   Relevanssi
  requête                        réf.        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
  requête                        réf.        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

ANALYSE PAR CONFIGURATION

5.1

WOOCOMMERCE NATIF

Vingt requêtes sur vingt-six produisent un ensemble vide. Les six qui produisent des résultats sont, sans exception, celles dont le terme recherché figure littéralement dans le titre du produit.

La recherche native interroge les champs `post_title` et `post_excerpt`. Elle n'accède ni aux références commerciales, ni aux attributs de déclinaison, ni aux taxonomies. L'hypothèse H1 est donc vérifiée : le défaut est de domaine et non de paramétrage, et aucune configuration de WooCommerce ne le corrige.

On relève en outre une latence de 3,85 secondes pour le rendu complet de la page de résultats, dont la composante imputable à l'interrogation de la base de données constitue une fraction minoritaire.

5.2

RELEVANSSI 4.28.2

Relevanssi construit un index inversé dans une table dédiée de la base de données du site, y intégrant attributs, taxonomies et champs personnalisés. Le défaut de couverture se réduit de 20 à 2 ensembles vides ; le rappel moyen passe de 19 % à 52 %. L'hypothèse H2 est vérifiée dans sa première partie.

Les coûts introduits sont de trois ordres.

LATENCE. Médiane 7 758 ms, maximum 33 643 ms, 19 cas sur 26 au-delà de 5 secondes. L'architecture prévoit deux requêtes SQL par terme — l'une pour l'ensemble des documents, l'autre pour la fréquence documentaire — avec recombinaison dans l'espace applicatif. Un syntagme de quatre termes entraîne huit requêtes.

L'indépendance de la latence vis-à-vis des ressources disponibles a été vérifiée : en triplant le buffer pool InnoDB de 128 à 384 Mo, la médiane diminue de 16 % (de 9 168 à 7 758 ms) et le maximum demeure inchangé. Le goulot d'étranglement n'est pas l'entrée-sortie.

OCCUPATION. L'indexation de 187 584 documents requiert 8 heures et 40 minutes et produit 1 598 117 lignes pour 512 Mo, situés dans la base de données applicative. La table obtenue est la deuxième du schéma par la taille, dépassée uniquement par `wp_postmeta` et équivalant à cinq fois `wp_posts`. Les index de la table (275 Mo) excèdent les données (237 Mo), conséquence de la nécessité de prendre également en charge la correspondance suffixale au moyen d'une colonne de termes inversés.

DÉGRADATION DE LA PRÉCISION. C'est le phénomène le plus notable. Relevanssi prévoit un mécanisme de repli qui, en l'absence de résultats sous conjonction, répète la requête en disjonction sans signalement à l'application appelante. L'effet sur la famille conversationnelle est documenté au tableau de la section 4 : quatre cas sur huit présentent une précision@100 n'excédant pas 0,02 sur des ensembles dont la cardinalité est comprise entre 983 et 1 471.

La portée du phénomène est méthodologique avant d'être pratique : sous une métrique de rappel seul, ces quatre cas figureraient parmi les meilleurs de la série.

LIMITES SUPPLÉMENTAIRES. Relevées par inspection du code et vérifiées sur les données :

```
troncature       un paramètre actif par défaut limite à 500 les documents
                   restitués par terme, ordonnés par fréquence brute avant
                   application du score de pertinence. Pour un terme de fréquence
                   documentaire 16 163, la fraction examinée est de 3 %
```

```
conjonction      l'opérateur ET n'est pas exprimé comme prédicat SQL mais
                   appliqué dans l'espace applicatif sur les listes déjà tronquées
```

```
morphologie      absence de racinisation pour la langue italienne. Vérifié :
                   giacca 16 470 / giacche 2 370 ; pantalone 17 368 /
                   pantaloni 2 373 ; camicia 9 221 / camicie 1 149 constituent
                   des termes distincts et non corrélés
```

```
seuil lexical    le paramètre de longueur minimale, fixé à 2, exclut de la
                   tokenisation les chaînes d'un caractère. Vérifié par comptage :
                   xs 5 537 — xl 5 834 — xxl 5 801 — 3xl 5 690 ; s, m, l absents
                   de l'index. La requête « taglia M » est donc insatisfiable par
                   construction
```

```
déclinaisons     les 167 584 documents de déclinaison constituent 89 % de
                   l'index, environ sept des 8 h 40 de traitement et 450 des
                   512 Mo. Le document restitué est la déclinaison, qu'une requête
                   de catalogue filtrée sur `post_type = product` exclut. La
                   requête sur référence de déclinaison produit un ensemble vide
```

MODÈLE DE PERTINENCE. Relevanssi adopte TF-IDF avec élévation au carré du facteur de fréquence documentaire inverse. Valeurs mesurées sur la population :

```
terme                df        IDF Relevanssi    IDF BM25
    composizione     19 500             5,12         2,26
    polo              4 559            13,82         3,72
    référence             1           131,08        11,74
    rapport rare/commun                25,6×         5,2×
```

L'amplification confère aux termes de faible fréquence documentaire un poids relatif quintuple de celui de la formulation BM25. Le comportement est optimal sur les références et sous-optimal sur les syntagmes multi-attributs, dans lesquels les termes devraient concourir de manière comparable.

Il est toutefois nécessaire de préciser que la substitution de la fonction d'ordonnancement ne serait pas résolutive. La normalisation par longueur documentaire prévue par BM25 favorise les documents brefs ; dans la population considérée, les documents brefs sont les déclinaisons, dont la longueur moyenne est de 4,9 termes contre 39,1 pour les produits, avec une cardinalité huit fois supérieure. À correspondance égale, BM25 placerait systématiquement les déclinaisons au-dessus des produits.

La conclusion qui s'en dégage est d'ordre plus général : la difficulté ne réside pas dans la fonction d'ordonnancement mais dans le modèle de données. Un catalogue n'est pas un corpus textuel, et les attributs de déclinaison ne sont pas des termes : ce sont des variables catégorielles à domaine fini.

5.3

MEILISEARCH 1.53.2

Meilisearch est un moteur de recherche libre implémenté en Rust, conçu pour être exécuté comme service indépendant de l'application. La différence architecturale avec une extension opérant sur la base de données applicative réside dans la séparation de la structure de recherche et de la structure de persistance : un index inversé optimisé pour l'accès par terme, en lieu et place d'un ensemble de tables relationnelles interrogées par correspondance textuelle.

La latence médiane mesurée est de 87 ms, dont 71 imputables à l'interconnexion réseau. Le temps de traitement déclaré par le moteur est compris entre 1 et 8 ms.

Les fonctionnalités pertinentes pour l'évaluation comprennent : tolérance aux fautes de frappe paramétrable par champ ; prédicats de filtrage sur attributs structurés ; distribution des facettes calculée conjointement à la requête ; critères d'ordonnancement multiples ; dictionnaires de synonymes et de mots vides ; règles de mise en avant conditionnées à la requête ; recherche par similarité sémantique avec inférence locale.

5.4

CELERSEARCH 1.7.2

CelerSearch est une extension WordPress qui achemine la recherche vers une instance Meilisearch externe. Elle n'inclut pas le moteur.

La configuration par paramètres prédéfinis obtient 10 résultats ok, 12 partiels et 4 ensembles vides, avec un rappel moyen de 66 % et une précision@100 de 64 %. La comparaison avec Relevanssi est nettement favorable sur l'ensemble des indicateurs : division par deux des ensembles vides, gain de 14 points de rappel, réduction de la latence médiane d'un facteur 89.

L'interface de configuration expose l'ensemble complet des paramètres de l'index : règles d'ordonnancement, attributs interrogeables, filtrables et triables, synonymes, mots vides, tolérance aux fautes, proximité, fragmentation des documents, limites de pagination, règles de mise en avant, recherche sémantique.

Deux difficultés de nature non paramétrique ont été relevées.

EFFICACITÉ D'INDEXATION. La transmission des documents s'effectue avec une cardinalité unitaire par requête HTTP (vérifié : `receivedDocuments = 1` sur toutes les opérations observées). Meilisearch est optimisé pour l'ingestion par lots. L'écart mesuré à moteur et population identiques :

```
lots de 2 000 documents       24 secondes pour 20 000 documents
    cardinalité unitaire          2,4 heures pour 20 000 documents
```

Facteur 354. Le paramètre de taille de lot exposé par l'extension régit la cardinalité des publications chargées par cycle applicatif, non celle des documents par requête. Ce choix est motivé dans le code par la volonté de maintenir un chemin d'écriture unique pour toutes les opérations — synchronisation en temps réel, file différée,

reconstruction complète — dont chacune appose un horodatage utilisé pour la suppression ultérieure des documents orphelins. La motivation est techniquement fondée ; le coût sur des populations de cette cardinalité est élevé.

Il est en outre relevé que le cycle d'indexation est piloté par le navigateur de l'opérateur au moyen d'appels asynchrones : l'interruption de la session interrompt le traitement.

RENSEIGNEMENT DU CHAMP DÉCLINAISONS. Le document produit prévoit un champ destiné aux références de déclinaison. Après indexation complète, ce champ est vide sur l'ensemble des documents examinés, bien qu'un compteur de déclinaisons soit renseigné. Vérification sur trois références de déclinaison d'un même produit : deux produisent un ensemble vide ; la troisième produit un résultat par coïncidence lexicale, la dénomination de la couleur figurant dans le titre du produit.

---

06 / BENCHMARK

EXPÉRIENCE À VARIABLE UNIQUE

La comparaison entre les deux configurations Meilisearch constitue le cœur de l'étude, en ce qu'elle permet d'isoler la contribution de la modélisation des données de celle de l'outil.

```
CONTRÔLÉES      moteur (Meilisearch 1.53.2), hôte, population (20 000
                  produits), ensemble des requêtes, ensemble de référence,
                  fenêtre temporelle, conditions réseau
```

```
MANIPULÉE       structure du document indexé et paramétrage de l'index
```

```
CelerSearch      configuration analytique
    ensembles vides             4                    0
    rappel moyen               66 %                 94 %
    précision@100              64 %                 78 %
    latence médiane            87 ms                87 ms
    indexation               2,4 heures           3 minutes
```

L'invariance de la latence confirme l'identité du moteur et la validité du contrôle. L'écart sur le rappel et la précision est donc intégralement imputable à la variable manipulée. L'hypothèse H3 est vérifiée.

On procède à la décomposition de l'écart en ses composantes.

---

07 / BENCHMARK

DÉCOMPOSITION DE L'ÉCART

7.1

DICTIONNAIRE DES MOTS VIDES

Les quatre ensembles vides de la configuration par défaut appartiennent tous à la famille conversationnelle :

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

En l'absence de dictionnaire de mots vides et sous stratégie de correspondance conjonctive, les prépositions et les termes de remplissage entrent dans l'ensemble des termes obligatoires. La probabilité qu'un document de catalogue contienne la préposition « da » dans son texte indexé est proche de zéro ; l'ensemble résultant est vide par construction.

Le dictionnaire retenu dans la configuration analytique comporte 60 formes italiennes. Le paramètre est exposé par l'interface de CelerSearch et vide par défaut.

7.2

DICTIONNAIRE DES SYNONYMES

La population ne contient aucune occurrence du lemme « vestito » : la dénomination retenue est « abito ». Une requête comportant le premier terme n'est pas satisfiable en l'absence d'une relation d'équivalence déclarée.

Effet mesuré sur la requête « vestito di seta rosso da donna », avec 13 relations d'équivalence déclarées : d'ensemble vide à 7 éléments sur 7 attendus, précision@100 unitaire.

7.3

EXPOSITION TEXTUELLE DES ATTRIBUTS

La requête « taglia 40 » comporte un terme — l'étiquette de l'attribut — qui ne figure dans aucun document. La valeur est présente comme modalité d'une variable catégorielle ; l'étiquette appartient au schéma, non aux données.

La configuration analytique génère lors de l'indexation un champ textuel associant étiquette et valeur :

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

Coût de stockage : quelques centaines d'octets par document. Effet mesuré sur la requête « jeans uomo taglia 40 » : de 17 éléments sur 71 attendus (rappel 24 %) à 71 sur 71 (rappel 100 %), avec précision@100 unitaire dans les deux cas.

7.4

AGRÉGATION DES RÉFÉRENCES DE DÉCLINAISON

La référence portée sur l'étiquette physique du vêtement est celle de la déclinaison, non du produit. Elle constitue la clé de recherche employée par l'acquéreur en possession de l'article, par l'opérateur logistique et par le revendeur.

Les configurations qui indexent la déclinaison comme document autonome produisent des ensembles vides sur les requêtes de catalogue, filtrées par type de publication. Les configurations qui ne l'indexent pas produisent le même résultat.

La configuration analytique agrège les 167 584 références de déclinaison dans les 20 000 documents de produit qui les contiennent. L'effet secondaire est notable : la cardinalité de l'index passe de 187 584 à 20 000 documents, et le temps de construction de 8 heures et 40 minutes à 3 minutes.

7.5

REPRÉSENTATION DES COMBINAISONS EXISTANTES

C'est la composante présentant le plus grand intérêt analytique et la moins fréquemment relevée.

L'application de deux prédicats de filtrage sur attributs multivalués — couleur et taille — produit, en l'absence d'une représentation de la combinaison, un ensemble comprenant les documents qui satisfont les deux prédicats sur des éléments distincts des tableaux respectifs.

Formellement : étant donné un produit p doté des ensembles de couleurs C(p) et de tailles T(p), et étant donné l'ensemble V(p) des déclinaisons effectivement existantes, le prédicat `c ∈ C(p) ∧ t ∈ T(p)` est une condition nécessaire mais non suffisante pour `(c,t) ∈ π(V(p))`.

L'écart est mesurable :

prédicat	filtres indépendants	combinaison	différence
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

total	1 476	1 402	74 (5,0 %)

5,0 % des résultats restitués par un filtre combiné ne correspondent pas à une déclinaison disponible. L'effet sur l'expérience d'achat est la sélection d'une combinaison épuisée.

La correction requiert la génération, lors de l'indexation, d'une clé composée pour chaque déclinaison existante, et son emploi comme prédicat de filtrage en lieu et place de la conjonction des prédicats isolés. Il ne s'agit pas d'un paramètre configurable : c'est une décision de modélisation qui présuppose l'analyse de la distribution des déclinaisons dans la population.

7.6

ORDONNANCEMENT DES CHAMPS INTERROGEABLES

Le cas est de faible ampleur et de grande valeur illustrative.

La requête « lino » produit, dans les deux configurations, des ensembles de cardinalité identique (2 803 éléments). La précision@100 diffère de 30 points : 70 % contre 100 %.

La cause réside dans l'ordre déclaré des attributs interrogeables, qui régit la règle d'ordonnancement par attribut. Dans la configuration par défaut, la dénomination du fournisseur réside dans le même conteneur que les attributs de produit et concourt à poids équivalent ; les produits du fournisseur dénommé « Cotone e Lino Sud » entrent en concurrence au classement avec ceux effectivement réalisés en lin. Dans la configuration analytique, le fournisseur constitue un champ distinct, placé après les attributs dans l'ordre de priorité.

---

08 / BENCHMARK

RECHERCHE PAR SIMILARITÉ SÉMANTIQUE

Il existe une classe de requêtes pour lesquelles aucune configuration lexicale ne peut produire de résultats, les termes employés n'appartenant pas au vocabulaire du catalogue :

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

L'ensemble de référence n'est pas vide : la population contient des manteaux en laine mérinos, des robes de cérémonie et des articles techniques. Les termes « caldo », « matrimonio », « correre » ne figurent cependant dans aucun document.

La recherche par similarité sémantique opère sur des représentations vectorielles denses. Chaque document et chaque requête sont projetés dans un espace métrique ; le critère d'ordonnancement est la proximité dans cet espace. La relation entre « matrimonio » et « abito da cerimonia » est encodée dans la géométrie de l'espace, non dans le recouvrement lexical.

Meilisearch incorpore le moteur d'inférence dans son exécutable et prend en charge l'exécution locale du modèle, sans dépendance à des services externes, clés d'accès ou coûts par requête. Les alternatives prises en charge comprennent les fournisseurs externes, les serveurs d'inférence propres et les vecteurs calculés en amont.

MESURES. Génération des représentations pour 20 000 documents : 85 minutes sur un cœur unique. Augmentation de l'index : de 162,5 à 263,4 Mo. Latence de requête en mode hybride : 42 à 51 ms.

Les cinq requêtes produisent des résultats pertinents : « cosa metto a un matrimonio » restitue des documents de la catégorie Abiti ; « qualcosa di caldo per l'inverno » restitue des articles en polaire et à dénomination thermique ; « vestiti per andare a correre » restitue des leggings.

DIFFICULTÉS DE CONFIGURATION RELEVÉES. Quatre, dont aucune n'est documentée en position apparente :

```
a) le modèle par défaut est entraîné sur corpus anglophone. Son emploi sur
     catalogue italien produit des représentations non discriminantes, sans
     aucun signalement
```

```
b) les modèles non prédéfinis exigent la spécification explicite de
     l'identifiant de révision ; à défaut, l'opération échoue avec une erreur de
     récupération de ressource non diagnostique de la cause
```

```
c) la classe d'entraînement du modèle est déterminante. Les modèles entraînés
     sur la similarité entre phrases produisent des résultats non pertinents là où
     sont requis des modèles entraînés sur la recherche asymétrique
     question-document
```

```
d) les dispositions de formulation documentées pour certains modèles, appliquées
     conformément à la spécification, dégradent les résultats et compromettent la
     composante lexicale de la recherche hybride
```

09 / BENCHMARK

MISE EN AVANT CONDITIONNELLE DES PRODUITS

Un moteur externe permet la définition de règles de mise en avant conditionnées à la requête : l'attribution d'une position fixe à des documents sélectionnés en présence de termes déterminés.

Mesure. Sur la requête « cappotto », l'ordonnancement par pertinence textuelle seule plaçait en première position un article ayant enregistré 26 ventes. L'application d'une règle de mise en avant sur le produit au plus fort volume place en première position l'article à 1 358 ventes.

AVERTISSEMENT MÉTHODOLOGIQUE. Les données de vente, d'évaluation et de remise n'étaient pas présentes dans la population générée et ont été simulées suivant une distribution en longue traîne, déterministe et reproductible, à seule fin de rendre la fonction démontrable. Les valeurs citées dans cette section sont vraisemblables mais non réelles.

Il est signalé que cette fonction, dans l'implémentation offerte par l'extension gratuite, n'était pas opérationnelle sur la version courante du moteur : le panneau permettait la composition de la règle mais l'opération d'enregistrement échouait, en raison de quatre incompatibilités de schéma entre la spécification implémentée et la spécification courante (dénomination d'un champ, structure du conteneur des conditions, forme d'une condition, base de l'indexation positionnelle). La composition manuelle de la règle dans la forme courante produit le comportement attendu.

Cette observation a une portée générale : la dépendance à un composant tiers pour une fonction à impact commercial direct comporte un risque de désalignement entre les cycles de publication du moteur et ceux du composant.

10 / BENCHMARK

LIMITES DE L'ÉTUDE

Elles sont explicitées car elles affectent la généralisabilité des conclusions et opèrent dans un sens défavorable aux thèses soutenues.

10.1

NATURE SYNTHÉTIQUE DE LA POPULATION

Les produits sont vraisemblables dans leur structure — taxonomie, attributs, déclinaisons, références — mais non dans leur contenu éditorial. Les descriptions sont générées à partir d'un vocabulaire clos : sur 187 920 termes distincts présents dans l'index Relevanssi, 187 584 ne figurent que dans un seul document (les références), laissant un vocabulaire partagé de 336 formes.

Cette limitation affecte de manière déterminante la section 8. Les représentations vectorielles sont calculées sur un texte largement répétitif, avec pour conséquence une compression de l'espace métrique : l'écart de score entre la première et la cinq-millième position s'est établi à 0,013. Les mesures de recherche sémantique doivent donc être interprétées comme démonstration du fonctionnement du mécanisme et non comme estimation de la qualité atteignable sur population réelle.

Il est en outre relevé que la génération combinatoire produit des associations incohérentes sur le plan marchand : 12 % des produits présentent des matières incompatibles avec la typologie (177 ceintures en cachemire, 165 en denim, 161 en polaire). Un système de recherche sémantique reflète fidèlement ces associations.

10.2

SOUS-ESTIMATION DE LA PRÉCISION

L'ensemble de référence est défini par correspondance textuelle sur le titre conjointe à l'appartenance taxonomique. Ce critère a été formulé lorsque le seul système évalué opérait par correspondance sur le titre, et pénalise les systèmes qui accèdent à la structure du catalogue.

Cas documenté : la requête « pantalone chino cotone » produit 227 éléments classés avec une précision@100 de 0,10. L'examen de l'ensemble restitué révèle 205 articles de typologie Chino, 17 Pantalone et 5 Chinos, tous appartenant à la catégorie Pantaloni et de matière Cotone ou Cotone Organico. Les résultats sont pertinents ; l'ensemble de référence exigeait la coprésence des deux lemmes dans le titre et excluait la modalité « Cotone Organico ».

Le critère n'a pas été modifié après observation des résultats. La révision d'un critère d'évaluation postérieurement à l'observation des résultats constitue un ajustement du critère au système mesuré.

10.3

SIMULATION DES SIGNAUX COMMERCIAUX

Comme indiqué en section 9.

10.4

ABSENCE D'ÉPREUVES DE CHARGE

Les relevés ont été conduits en condition d'utilisateur unique. Le comportement sous concurrence, sur un hôte à cœur unique, n'a pas été évalué.

10.5

RÉPLICATION UNIQUE

La variance entre exécutions répétées sur infrastructures distinctes n'a pas été estimée. Les latences rapportées sont des médianes sur série unique.

10.6

DÉFAUT CONNU DE LA CONFIGURATION ANALYTIQUE

La requête « taglia M » produit 16 189 éléments pour 4 258 attendus. La cause est la correspondance préfixale appliquée au dernier terme de la requête : le caractère « M » isolé correspond à toute forme l'ayant pour préfixe, condition satisfaite par l'ensemble des documents de la population. L'ordonnancement atténue l'effet — les vingt premiers éléments portent tous la modalité M — mais la cardinalité annoncée est erronée. La correction requiert la reconnaissance des valeurs de taille lors de l'interprétation de la requête et leur traduction en prédicat de filtrage.

---

11 / BENCHMARK

CONCLUSIONS

11.1 L'hypothèse H1 est vérifiée. La recherche native de WooCommerce présente un

> défaut de domaine : elle n'accède ni aux attributs de déclinaison, ni aux taxonomies, ni aux références commerciales. Sur une population comportant des déclinaisons, il en résulte l'insatisfiabilité de la majorité des requêtes commercialement pertinentes — vingt sur vingt-six dans le présent relevé. Le défaut n'est pas corrigible par paramétrage.

11.2 L'hypothèse H2 est vérifiée. Les extensions opérant sur la base de données

> applicative réduisent le défaut de couverture (de 20 à 2 ensembles vides) en introduisant des coûts en latence (médiane 7 758 ms) et en occupation (512 Mo dans la base du site). Un mécanisme de repli non signalé est en outre relevé, produisant des ensembles de forte cardinalité et de précision proche de zéro, classifiables comme succès par des métriques de rappel seul.

11.3 La substitution du moteur résout simultanément la latence et l'occupation,

```
avec des écarts d'ordre de grandeur : 87 ms contre 7 758, index externe à la
      base applicative, construction en 3 minutes contre 8 heures et 40.
```

11.4 L'hypothèse H3 est vérifiée, et constitue le résultat principal. À moteur,

```
hôte, population et requêtes identiques, l'écart entre configuration par
      défaut et configuration analytique est de 28 points de rappel et 4 ensembles
      vides. Cet écart est intégralement imputable à six décisions de modélisation
      et de paramétrage, documentées en section 7 et quantifiées individuellement.
```

11.5 Aucune de ces six décisions n'est déterminable automatiquement, chacune

```
dépendant de propriétés spécifiques de la population : langue, conventions
      de dénomination adoptées, axes de déclinaison employés, lexique du secteur
      marchand, distribution des déclinaisons.
```

Il s'en déduit que l'adoption d'un moteur de recherche externe constitue une condition nécessaire mais non suffisante. La composante résiduelle — quantifiée à 28 points de rappel, correspondant à la classe de requêtes à plus forte intention d'achat — requiert une analyse préalable de la population et une conception explicite du schéma documentaire.

ANNEXE

ANNEXE A — REPRODUCTIBILITÉ

La population est générée avec des paramètres déterministes (graine 20260906) et reconstructible à l'identique. L'ensemble des 26 requêtes, les critères de construction de l'ensemble de référence et les deux métriques sont documentés aux sections 2.3 à 2.5 et applicables à tout catalogue WooCommerce.

Versions employées : 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.

ANNEXE

ANNEXE B — NOTATION

```
R              ensemble de référence pour une requête donnée
  P              ensemble restitué par le système
  P₁₀₀           cent premiers éléments de P selon l'ordonnancement du système
  rappel         |R ∩ P| / |R|
  précision@100  |R ∩ P₁₀₀| / min(100, |P|)
  df             fréquence documentaire d'un terme
  n              cardinalité de l'ensemble restitué
```

MI

Massimo Ivaldi

A SEARCH COMPANY

CONCLUSION

La recherche n’est pas une case.

C’est un système.

Le moteur compte. Mais le résultat se construit en amont : dans la structure du catalogue, le choix des champs, les synonymes, les déclinaisons et la manière dont l’intention réelle des utilisateurs est traduite en données.

01 / Mesurer

Partir de requêtes réelles, du rappel et de la précision. Pas d’impressions.

02 / Modéliser

Placer les attributs, les déclinaisons et les combinaisons au bon endroit.

03 / Configurer

Traiter les synonymes, la tolérance et le classement comme des décisions métier.

04 / Vérifier

Rendre chaque choix reproductible, documenté et améliorable.

massimoivaldi.com · A search company

septembre 2026