MI

Massimo Ivaldi

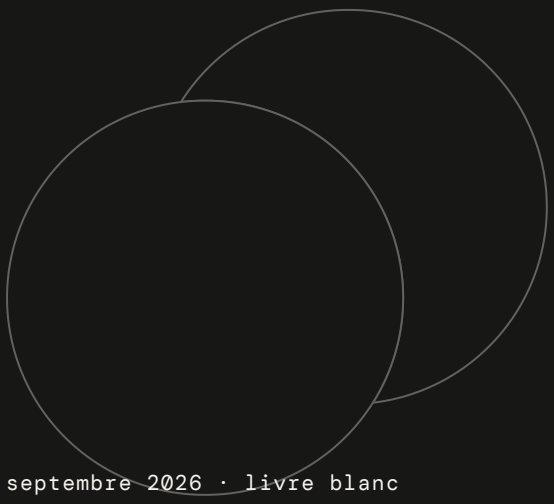
A SEARCH COMPANY

LIVRE BLANC / RECHERCHE SÉMANTIQUE · E-COMMERCE

# Recherche sémantique pour l'e-commerce, une étude de cas

## De 36 requêtes sans réponse à zéro

Données, méthode et limites d'une intervention réelle sur un catalogue technique.



septembre 2026 · livre blanc

LA RECHERCHE COMME SYSTÈME DE CROISSANCE.

# Des mots au sens.

Un cas réel portant sur 1 535 produits montre comment de meilleures données et une recherche hybride peuvent éliminer les requêtes sans réponse et réduire fortement les pires temps d'attente.

**1535**
PRODUITS DANS LE CATALOGUE

**121**
REQUÊTES DE TEST

**36 → 0**
REQUÊTES SANS RÉPONSE

**3 237 → 168**
MILLISECONDES DANS LE PIRE DES CAS

DANS CETTE ÉTUDE DE CAS

01 **Résumé**

02 **Qu'est-ce que la recherche sémantique, c'est-à-dire par le sens**

03 **La méthode: d'abord les données, ensuite la mesure**

04 **L'enrichissement du catalogue**

05 **Le jeu de test: comment on mesure une recherche**

06 **La mesure de départ: la recherche native de PrestaShop**

07 **Le passage à Meilisearch**

08 **Les synonymes: une vérification qui change l'approche**

09 **Le résultat**

10 **Ce que ces chiffres ne disent pas**

11 **En résumé**

**Le résultat central**

À catalogue identique et sur 121 requêtes, les recherches sans réponse passent de 36 à zéro. Le temps dans le pire des cas descend de 3 237 à 168 millisecondes.

RÉSUMÉ

# Résumé

La recherche interne d'une boutique en ligne est souvent l'endroit précis où l'on perd les clients qui avaient déjà décidé d'acheter. Ce document décrit une intervention complète sur un catalogue d'outillage de mille cinq cent trente-cinq articles, menée en deux temps: d'abord la reconstruction des données produit, ensuite le remplacement du moteur de recherche par un moteur capable de chercher par le sens et non seulement par correspondance de lettres.

Le travail a été mesuré avec un ensemble de cent vingt et une questions de test, en partie construites par degrés de difficulté croissante, en partie collectées à partir de ce que les gens tapent réellement dans les moteurs de recherche. Les mêmes questions ont été posées à l'ancien moteur et au nouveau, à catalogue identique.

Le résultat: les questions qui ne recevaient aucune réponse passent de trente-six à zéro, et le temps d'attente dans le pire des cas descend de trois secondes deux à cent soixante-huit millisecondes. Le document rapporte également ce que la mesure ne démontre pas, et les points où la recherche sémantique, utilisée sans précautions, produit des erreurs pires que celles qu'elle résout.

L'intervention a été réalisée sur PrestaShop parce que c'était la plateforme du client. La méthode ne dépend pas de la plateforme: les mêmes étapes s'appliquent à WooCommerce, Magento, Shopify et à un site de vente développé sur mesure. Ce qui change, ce sont les noms des champs et la manière de les ajouter, pas la substance du travail.

Un de nos clients vend de l'outillage et du matériel de travail avec une boutique PrestaShop. Mille cinq cent trente-cinq articles: perceuses, visseuses, batteries, pinces, visserie, meuleuses. Un catalogue petit mais technique, où chaque pièce a des dimensions, des tensions, des matériaux et des compatibilités.

Le problème, tel qu'il nous a été exposé, est celui que nous entendons formuler presque toujours avec les mêmes mots: "la recherche du site ne fonctionne pas". Celui qui cherche ne trouve pas, et celui qui ne trouve pas s'en va.

Derrière cette phrase il y a toujours trois causes qui s'additionnent: un moteur de recherche construit pour retrouver ce dont on connaît déjà le nom, un catalogue rédigé dans la langue du fabricant plutôt que dans celle du client, et des questions qui décrivent un problème au lieu de nommer un produit. Les trois causes s'aggravent mutuellement, et aucune des trois ne se résout toute seule.

Ce document raconte ce que nous avons trouve, ce que nous avons fait, et de combien le résultat s'est déplace, avec les chiffres.

CASO STUDIO

# Qu'est-ce que la recherche sémantique, c'est-à-dire par le sens

Une recherche traditionnelle compare des lettres. Si vous écrivez "disqueuse", le mot que les gens du métier emploient pour désigner une meuleuse d'angle, et que ce mot ne figure pas dans le catalogue, elle ne trouve rien, même si la boutique est pleine de meuleuses.

La recherche par le sens fonctionne autrement. Chaque fiche produit est transformée en une longue liste de nombres, appelée **vecteur**, qui représente le sens du texte. La question du client est transformée de la même manière. Le moteur cherche ensuite les fiches dont le vecteur ressemble le plus à celui de la question. Le calcul qui

transforme un texte en nombres s'appelle **embedding**, et il est effectué par un modèle d'intelligence artificielle entraîné pour cela.

C'est ce qui permet à "un outil pour enlever la gaine des fils" d'aboutir à une pince à dénuder sans partager un seul mot avec la fiche.

## Ce que cela résout

L'avantage n'est pas que le moteur trouve davantage: c'est qu'il répond à celui qui **ne sait pas comment s'appelle la chose qu'il veut**. Celui qui connaît la référence de l'article la trouvait déjà avant, et celui qui connaît le nom technique aussi. La recherche sémantique sert à tous les autres, qui sont la majorité, et qui aujourd'hui repartent les mains vides.

Trois exemples tirés des mesures de ce travail, tous avec la recherche existante à zéro résultat.

**Celui qui décrit le problème au lieu de nommer l'objet.**

```
que faut-il pour percer le mur   ->   Perforateur AEG BH 24E SDS-plus
```

Dans la fiche de ce perforateur n'apparaissent ni "faut-il", ni "percer", ni "mur". Y apparaissent "perçage", "maçonnerie", "béton". Le modèle sait qu'il s'agit du même domaine de sens.

**Celui qui parle comme il parlerait à un vendeur.**

```
je voudrais une visseuse légère pour la maison   ->   Visseuse électrique AEG S 4000 E
```

**Celui qui emploie le nom du métier plutôt que celui du fabricant.**

```
un outil pour dénuder les fils   ->   Pince à dénuder Microconnect PP-LSA2
```

Le verbe employé par le client n'existe pas dans le catalogue: la fiche dit "pince à dénuder" et "retirer la gaine".

Sur ce catalogue, les questions de ce type représentaient environ un tiers du total, et la recherche existante n'en servait aucune.

## La où elle se trompe, et pourquoi elle doit être encadrée

Le vecteur ne sait pas qu'il ne sait pas.

Une recherche par mots, lorsqu'elle ne trouve pas, renvoie zéro résultat: c'est désagréable mais honnête, et le client comprend que cette chose-la n'y est pas. Une recherche par le sens renvoie **toujours** quelque chose, parce que parmi tous les produits il y en a toujours un plus proche que les autres. Parfois il est proche pour la mauvaise raison.

Trois cas, tous tirés des essais avec la recherche purement sémantique.

**Elle suit la sonorité du mot plutôt que le sens.**

```
pince coupante   ->   Scie à onglet AEG PS 216 L, 1.500 W
```

Le client cherchait une pince coupante de poche, on lui propose une machine d'établi de mille cinq cents watts. Dans la langue du catalogue les deux objets s'appellent "tronchese" et "troncatrice": ils n'ont rien en commun sinon les cinq premières lettres, et le modèle a suivi la sonorité plutôt que le sens.

**Elle se trompe de référence, et de peu.**

```
WZ0060   ->   Pince LogiLink WZ0068
```

C'est le cas le plus insidieux, parce que le résultat **paraît juste**: le client commande, reçoit la mauvaise pièce, et c'est la boutique qui en porte la responsabilité. Lors des essais, la recherche purement sémantique s'est trompée sur quatre références sur cinq.

**Elle associe de manière sensée et fausse.**

Sur un autre catalogue, en cherchant le nom populaire de l'interrupteur différentiel, elle a proposé quatre onduleurs. Le raisonnement se comprend, un appareil qui sauve le courant, mais le client cherchait le différentiel du tableau électrique.

D'ou une règle valable pour quiconque évalue un moteur de recherche: **le décompte du nombre de recherches qui donnent au moins un résultat n'est jamais une mesure suffisante.** Un moteur qui se trompe avec assurance gagne ce décompte contre un moteur qui admet ne pas savoir.

Et d'où la conséquence pratique, qui est la part du métier: la recherche sémantique ne se met pas en production seule. Elle se mélange à la recherche par mots, dans une proportion qu'il faut choisir et mesurer sur le catalogue de cette boutique-la.

## Deux installations sémantiques ne donnent pas le même résultat

"Recherche sémantique" est une étiquette qui recouvre des résultats très différents les uns des autres. Deux choses en particulier sont tenues pour acquises et ne le sont pas.

**Les embeddings ne se valent pas tous.** Les modèles qui transforment le texte en nombres sont nombreux, de fabricants différents, et ils diffèrent sur trois axes qui comptent tous.

| | |
|---|---|
| la langue | presque tous sont entraînés surtout sur l'anglais. Un modèle qui comprend bien "angle grinder" peut comprendre bien moins "meuleuse d'angle", et presque pas le mot du métier |
| le secteur | le langage de la quincaillerie, avec ses cosses, ses clips faston, ses embouts et ses têtes fraisées, est loin de la langue courante sur laquelle les modèles s'entrainent |
| coût et taille | un modèle plus grand produit des vecteurs plus longs, coûte plus cher à générer et occupe plus de place dans l'index, sans toujours mieux répondre |

Sur un catalogue technique et dans une langue autre que l'anglais, essayer deux ou trois modèles et les comparer sur les mêmes questions fait la différence entre une installation qui fonctionne et une installation qui en a l'air.

**Le modèle doit être alimenté correctement.** Un moteur sémantique vaut ce que vaut le modèle qu'il embarque et ce que vaut le texte qu'on lui remet: ce sont deux parties de la même installation, et elles se construisent ensemble. C'est la part invisible du travail, car celui qui regarde un site n'a aucun moyen de savoir ce qui a été donné au modèle au moment de la construction de l'index.

| CE QUE L'ON ENVOIE | CE QUI SE PASSE |
|---|---|
| seulement le titre | le vecteur représente peu de chose, et sur des titres faits de sigles il ne représente rien |
| tout le document, références et prix compris | le vecteur se dilué: une référence article n'a pas de sens à représenter, c'est du bruit |
| une sélection raisonnée | le vecteur représente vraiment le produit |

CASO STUDIO

# La méthode: d'abord les données, ensuite la mesure

La tentation, en arrivant, est de mesurer tout de suite. C'est une erreur, et pour une raison technique avant d'être méthodologique.

Mesurer un moteur de recherche sur un catalogue dont les titres ne disent pas ce qu'est le produit ne mesure pas le moteur: cela mesure le catalogue. Un moteur, aussi bon soit-il, ne peut pas trouver un mot qui n'est écrit nulle part dans le catalogue. En partant de données fausses, on obtiendrait de mauvais chiffres pour les deux moteurs, et on finirait par attribuer à l'un le défaut de l'autre.

Nous avons donc d'abord reconstruit les données, puis mesuré. Et nous avons mesuré les deux moteurs **à catalogue identique**, en donnant à chacun les mêmes informations, chacune écrite dans le champ que ce moteur sait lire: c'est la seule manière pour que la comparaison résiste à qui la conteste.

L'ordre des deux temps ne dit rien de leur poids respectif. La reconstruction des données est la condition qui permet au nouveau moteur de rendre ce qu'il peut rendre; mais à elle seule, comme on le verra dans les chiffres, elle récupère deux questions sur cent vingt et une, tandis que le changement de moteur les récupère toutes les trente-six.

CASO STUDIO

# L'enrichissement du catalogue

## Ce qui n'allait pas

| DÉFAUT | COMBIEN DE PRODUITS |
|---|---|
| titre qui ne nomme pas l'objet | 192 sur 1.535, soit 12,5% |
| description qui répète les données sans rien ajouter | environ 90% |
| description rédigée par un être humain | 157, soit 10% |

Un exemple de titre: **"AEG BBH 12"**. C'est un perforateur, mais le titre ne le dit pas, et le mot "perforateur" n'apparaît nulle part dans la fiche. Celui qui cherche "perforateur AEG" y arrive par hasard.

Les descriptions étaient des listes à puces générées à partir des données: "Dispositif à courant résiduel, Nombre de pôles: 1, IP20, Blanc". Elles n'ajoutent pas un mot à ceux qui figurent déjà dans les caractéristiques techniques.

## Comment nous l'avons corrigé

Nous avons donné à un modèle d'intelligence artificielle, produit par produit, **l'enregistrement complet de la fiche et la photographie**, en lui demandant trois choses: un titre sous la forme nom de l'objet, marque, modèle; une description longue et rédigée qui contienne toutes les caractéristiques techniques mais qui dise aussi à quoi sert l'objet, qui l'utilise et dans quels travaux; une description courte.

La photographie compte plus qu'on ne le croit. Sur une pince à dénuder, le modèle a écrit "présente une structure à pince avec des mâchoires réglables et une vis de réglage latérale": ce détail ne figurait pas dans l'enregistrement, il l'a vu sur l'image.

## Le point critique: le modèle inventé

C'est la partie qui distingue un travail bien fait d'un désastre.

Le texte des instructions disait clairement de n'inventer aucune donnée. **Le modèle a inventé quand même**: il a écrit quatorze codes-barres inexistants, treize chiffres, format correct, sur des produits qui n'ont pas de code-barres.

Cela ne se résout pas en réécrivant les instructions. Cela se résout par des contrôles automatiques en aval, avant d'écrire quoi que ce soit dans le catalogue. Les notres étaient au nombre de six: structure de la réponse valide, titre dans la longueur maximale et contenant marque et modèle, descriptions dans les limites, ponctuation normalisée, et **aucun nombre de treize chiffres qui ne soit le code-barres réel du produit**.

Ce qui ne passe pas est réessayé, et si cela ne passe toujours pas, c'est mis de côté et non écrit.

CASO STUDIO

# Le résultat

| | |
|---|---|
| fiches réécrites | 1.535 sur 1.535 |
| passages nécessaires | 4 |
| codes-barres inventés et bloqués | 14 |
| coût total du modèle | environ 1 dollar |
| vocabulaire de la recherche interne | de 9.001 à 16.055 mots |

Le vocabulaire presque double est la donnée la plus parlante: ce sont les mots qui n'existaient auparavant en aucun point du catalogue, et qui rendent désormais les produits trouvables.

CASO STUDIO

# Le jeu de test: comment on mesure une recherche

Pour affirmer qu'une recherche s'est améliorée il faut des questions, et ces questions ne peuvent pas être inventées par celui qui fait la mesure, sinon on finit par écrire des requêtes dans la langue du catalogue, qui fonctionnent toujours.

Nous avons construit le jeu de test à partir de deux sources.

**Cinquante-cinq requêtes construites sur dix degrés de difficulté croissante**, du nom seul à la phrase parlée:

| DEGRÉ | EXEMPLE |
|---|---|
| 1. le nom de la chose | `perceuse` |
| 2. nom et marque | `perforateur AEG` |
| 3. nom et attribut | `perceuse à percussion` |
| 4. attribut numérique | `batterie 18v` |
| 5. référence | `BBH18-LI402C` |
| 6. nom populaire | `disqueuse , clé Allen` |
| 7. le but au lieu de l'objet | `outil pour dénuder les fils` |
| 8. catégorie large | `outillage électroportatif` |
| 9. plusieurs contraintes | `perceuse sans fil 18v légère` |
| 10. phrase parlée | `il me faut une perceuse pour percer le béton, sans fil` |

**Une cinquantaine de phrases réelles**, collectées auprès de l'autocompletion de Google. Le procédé: on prend les noms des catégories et des marques du catalogue, on les soumet à l'autocompletion, et reviennent les phrases que les gens tapent réellement. A partir de 253 amorces sont sorties 1.407 suggestions.

Au total **121 requêtes distinctes**.

## Le premier résultat arrive avant la mesure

Les phrases réelles échouaient deux fois plus que celles que nous avions construites nous-mêmes:

| ORIGINE DES REQUÊTES | COMBIEN NE TROUVENT RIEN |
|---|---|
| construites à la table | 18% |
| collectées sur Google | 38% |

La raison est instructive. Les notres étaient écrites avec les mots du catalogue. Les vraies sont écrites avec les mots de ceux qui travaillent: **cosses, faston, embouts, à cliquet, autoforeuses, tête fraisée, sds max, orbitale, decaper**. Des mots que le catalogue ne contient pas, et là où le mot manque, le résultat manque.

Celui qui mesure sa propre recherche avec des requêtes qu'il a inventées obtient toujours une bonne note.

---

CASO STUDIO

# La mesure de départ: la recherche native de PrestaShop

PrestaShop construit un **index inversé**, c'est-à-dire une liste qui, pour un mot donné, indique immédiatement dans quels produits il apparaît. A chaque mot il attribue un poids selon le champ d'où il provient: la référence article pèse 10, le nom du produit 6, les étiquettes 4, la catégorie et la marque 3, les caractéristiques 2, les descriptions 1.

Lorsqu'une question arrive, il la découpe en mots, écarte ceux de moins de trois caractères, cherche chaque mot **par préfixe** (celui qui cherche "nastr" trouve "nastro"), et ne conserve que les produits qui contiennent tous les mots. Puis il classe par somme des poids.

Il dispose aussi d'un rattrapage des fautes de frappe: si un mot ne trouve rien, il cherche dans l'index celui qui lui ressemble le plus. Cela fonctionne, mais cela coûte.

| | |
|---|---|
| requêtes sans aucun résultat | 38 sur 121 |
| temps médian | 20 ms |
| temps dans le pire des cas | 2.292 ms |

Le pire temps mérite une ligne: **les plus mauvaises réponses sont aussi les plus lentes**. Lorsque le moteur ne trouve pas le mot, il se met à chercher le plus ressemblant dans tout le vocabulaire, y passe deux secondes, et au bout du compte ne trouve rien quand même. Le client paie l'attente la plus longue pour la plus mauvaise réponse.

---

CASO STUDIO

# Le passage à Meilisearch

Meilisearch est un moteur de recherche qui vient s'adjoindre à la boutique: le catalogue lui est remis sous forme de documents, un par produit, et il les indexe.

## Les graduations: la recherche hybride

Meilisearch peut travailler de deux manières simultanément, en cherchant les mots et en cherchant le sens, et **un seul paramètre décidé du poids de l'un et de l'autre**. Il s'appelle `semanticRatio` et c'est un nombre compris entre 0 et 1.

| VALEUR | COMPORTEMENT |
|---|---|
| 0 | seule compte la recherche par mots |
| 0,3 | trois dixièmes du score viennent du sens, sept des mots |
| 0,5 | moitié-moitié |
| 0,7 | sept dixièmes du sens, trois des mots |
| 1 | seul compte le sens, les mots sont ignorés |

Ce n'est pas un bouton de volume: c'est une répartition, et les deux parts doivent faire un. Il n'existe pas de valeurs supérieures à 1, et le moteur les refuse.

## Ce que nous avons envoyé au modèle d'embedding

C'est la décision qui pèse le plus et que personne ne voit.

Ce n'est pas le document entier qui va au modèle. C'est **un seul champ**, composé par nos soins, qui contient dans l'ordre: titre, marque, chemin du rayon, noms populaires de l'objet, description courte, description longue, caractéristiques techniques sous forme lisible.

Restent en dehors les références, le prix, le stock, les adresses des images et les données d'emballage. Une référence article n'a pas de sens à représenter: ce serait du bruit qui déplace le vecteur sans rien ajouter. Ces champs restent dans le document, cherchables par mots et utilisables dans les filtres, ils n'entrent simplement pas dans le vecteur.

Il y a aussi une limite technique à connaître: Meilisearch tronque le texte destiné au modèle à un seuil configurable. Avec la valeur que nous avions choisie au départ, **un produit sur cinq était coupe**, et la partie coupée était les caractéristiques techniques, qui se trouvent à la fin. C'eût été un défaut invisible: l'index se construit, tout paraît en ordre, et un cinquième du catalogue a un vecteur mutilé. Le seuil relevé, plus aucun produit n'est tronque.

## Coûts et temps de construction

| | |
|---|---|
| documents charges | 1.535 |
| temps de chargement, vecteurs compris | 15 secondes |
| coût des embeddings | environ 1,5 centime |

---

CASO STUDIO

# Les synonymes: une vérification qui change l'approche

Face à "disqueuse" qui ne trouve pas les meuleuses, la réponse évidente est: mettons un synonyme dans le moteur. Tous les moteurs de recherche disposent d'une table de synonymes, où l'on déclare que deux mots doivent être considérés comme équivalents.

Cela fonctionne, mais pas sur tous les types de recherche. Nous l'avons vérifié avec un essai construit de telle sorte que la seule chose à changer soit la présence du synonyme: un index séparé avec six produits, trois meuleuses d'angle et trois perceuses, interrogé quatre fois avec le même mot.

| TYPE DE RECHERCHE | SYNONYME | PREMIER RÉSULTAT | SCORE |
|---|---|---|---|
| mots seuls | absent | **aucun résultat** | |
| mots seuls | actif | Meuleuse d'angle | |
| sens seul | absent | Meuleuse d'angle AEG | 0,0157 |
| sens seul | actif | Meuleuse d'angle AEG | 0,0157 |

Les deux premières lignes disent que **le synonyme fonctionne**: sur la recherche par mots, sans lui le mot "disqueuse" ne trouve rien, avec lui les meuleuses sortent.

Les deux dernières disent que **sur la recherche par le sens il ne fait absolument rien**: avec et sans le synonyme, le résultat est identique, et identique jusqu'au score. Ce score est un nombre calculé par le moteur: si le synonyme avait eu le moindre effet, il aurait change. Il n'a pas change d'un dix-millième.

Que dans ces deux lignes sorte malgré tout une meuleuse n'est pas le mérite du synonyme: cela tient au fait que dans un index de six produits seulement, le vecteur du mot finit de toute façon plus près d'une meuleuse que d'une perceuse. Sur le catalogue entier, mille cinq cents produits, la recherche par le sens seul, interrogée avec ce mot, ne trouve pas les meuleuses du tout.

## Pourquoi

Le vecteur du produit est calculé **au moment du chargement du catalogue**, à partir du texte de la fiche. Aucune table de substitution ne peut modifier un nombre déjà écrit dans l'index. Et la question du client est transformée en vecteur telle quelle, sans passer par les synonymes.

Les synonymes du moteur fonctionnent donc à cent pour cent lorsque l'on cherche par mots seuls, à moitié dans la recherche hybride, et pas du tout dans la recherche purement sémantique.

## La voie alternative, et pourquoi elle est meilleure

Si en revanche on écrit le mot **dans le catalogue**, il fonctionne sur les deux canaux. Et il en gagne un troisième.

| OÙ L'ON MET LE MOT | RECHERCHE PAR MOTS | RECHERCHE PAR LE SENS | GOOGLE |
|---|---|---|---|
| table de synonymes du moteur | oui | non | non |
| dans la fiche produit | oui | oui | oui |

La table de synonymes n'est lue que par le moteur de recherche interne. Le mot écrit dans la fiche est lu aussi par Google. Un produit décrit comme "dispositif à courant résiduel" ne se trouve pas en cherchant "différentiel", ni dans le site ni sur les moteurs externes.

**Le mot écrit dans le catalogue se paie une fois et travaille sur deux canaux: dans le moteur de recherche de la boutique et au-dehors, sur les moteurs de recherche externes.** C'est pourquoi nous le préférons à la table de synonymes, qui reste un outil du moteur et doit servir à ce qu'elle sait faire.

## Les noms du client

De là, la solution adoptée. Pour chacune des 75 catégories du catalogue nous avons défini les noms par lesquels les gens demandent cet objet, qui ne sont pas ceux par lesquels le fabricant le décrit:

| CATÉGORIE AU CATALOGUE | NOMS DU CLIENT |
|---|---|
| Meuleuses d'angle | disqueuse, meuleuse, touret |
| Perforateurs | perfo, marteau perforateur, burineur |
| Pinces à dénuder | dénude-fil, pince à dénuder, dégaineur |
| Clés six pans | clé Allen, clé BTR, clé en L |
| Agrafeuses manuelles | agrafeuse, cloueur à main, pistolet à agrafes |
| Visseuses électriques | visseuse, tournevis électrique |
| Pinces coupantes | pince coupante, coupe-câble, tenaille |

La donnée vit **à l'intérieur de la boutique**, et non dans un fichier externe: un module ajoute à la catégorie un champ dédié, que le visiteur du site ne voit pas et que la personne qui gère la boutique peut modifier. L'exportateur le relit depuis là et le porte à la fois dans le document et dans le texte à partir duquel le vecteur est calculé.

C'est la partie du travail qui se rentabilise le mieux, car elle vaut aussi pour le positionnement sur les moteurs de recherche externes.

Sur PrestaShop le champ s'ajoute avec un module; sur WooCommerce avec un champ personnalisé; sur Magento avec un attribut; sur un site sur mesure avec une colonne. Le principe ne change pas: **un attribut de la catégorie, à l'intérieur de la boutique, non exposé au client, exportable.**

---

CASO STUDIO

# Le résultat

Les mêmes 121 requêtes, le même catalogue, les mêmes noms du client donnés aux deux moteurs, chacun dans le champ que son programme sait lire.

| | PRESTASHOP NATIF | MEILISEARCH À 0,3 |
|---|---|---|
| requêtes sans aucun résultat | 36 | 0 |
| temps médian | 33 ms | 112 ms |
| temps dans le pire des cas | 3.237 ms | 168 ms |

Le temps médian se dégrade de quatre-vingt-dix millisecondes, et ces quatre-vingt-dix comprennent le trajet jusqu'au centre de données et retour, alors que PrestaShop tourne sur la machine même de la boutique.

En échange, **la longue traîne disparaît**: le pire des cas passe de trois secondes deux à cent soixante-huit millisecondes. Pour qui utilise un site, le plafond compte plus que la moyenne.

## Pourquoi 0,3 et pas une autre valeur

Nous avons mesuré les 121 requêtes à six valeurs différentes.

| VALEUR | REQUÊTES SANS RÉPONSE | LES CINQ RÉFÉRENCES DE TEST | TEMPS MAXIMAL |
|---|---|---|---|
| 0 (mots seuls) | 5 | 5 sur 5 | 442 ms |
| 0,2 | 0 | 5 sur 5 | 142 ms |
| 0,3 | 0 | 5 sur 5 | 168 ms |
| 0,5 | 0 | 5 sur 5 | 156 ms |
| 0,7 | 0 | 5 sur 5 | 150 ms |
| 1 (sens seul) | 0 | **1 sur 5** | 184 ms |

Vingt pour cent de sens suffisent à combler tous les trous. Les références articles tiennent jusqu'à 0,7 et ne s'effondrent qu'a l'extrême.

En observant de combien changent les premiers résultats en passant d'une valeur à la suivante, on découvre trois zones et non une échelle continue:

| PASSAGE | REQUÊTES OÙ LA TÊTE DU CLASSEMENT CHANGE |
|---|---|
| de 0 à 0,2 | 10 sur 121 |
| de 0,2 à 0,3 | 2 |
| de 0,3 à 0,5 | 20 |
| **de 0,5 à 0,7** | **92** |
| de 0,7 à 1 | 18 |

Jusqu'à 0,5 le moteur reste pour l'essentiel lexical et le sens se contente de combler les trous. Au-dessus de 0,5 c'est le sens qui commande. **0,3 est la valeur prudente**: elle comble tous les trous, ne touche pas aux références, et ne bouleverse pas le comportement auquel le client de la boutique s'attend.

## Quelques exemples concrets

| REQUÊTE | PRESTASHOP | MEILISEARCH À 0,3 |
|---|---|---|
| `disqueuse` | Aspirateur de chantier Rowenta | Meuleuse d'angle AEG WS 12-125 |

| | | |
|---|---|---|
| `perfo` | aucun résultat | Perforateur AEG BBH 12 |
| `outil pour dénuder les fils` | aucun résultat | Pince à dénuder Microconnect PP-LSA2 |
| `que faut-il pour percer le mur` | aucun résultat | Perforateur AEG BH 24E SDS-plus |
| `je voudrais une visseuse légère pour la maison` | aucun résultat | Visseuse électrique AEG S 4000 E |
| `pour faire briller le sol` | Cireuse Hoover F 3870 | Cireuse à parquet AEG PE 150 |
| `BBH18-LI402C` | Perforateur AEG BBH18-LI402C | Perforateur AEG BBH18-LI402C |

## Pourquoi les données seules ne suffisent pas

La reconstruction du catalogue améliore quelque chose même sur l'ancien moteur, et sur certains mots cela se voit.

| QUESTION | ANCIEN MOTEUR, AVANT | ANCIEN MOTEUR, APRÈS |
|---|---|---|
| `perfo` | aucun résultat | 30 résultats |
| `disqueuse` | 3 aspirateurs | 18 meuleuses |
| `tournevis` | 62 résultats | 114 résultats |

Mais sur l'ensemble des cent vingt et une questions, ce gain vaut **deux questions récupérées**: de trente-huit sans réponse à trente-six. Tout le reste, les trente-six autres, c'est le nouveau moteur qui le récupère.

La raison est structurelle et ne se contourne pas en rédigeant mieux les fiches. Un moteur qui compare des lettres ne peut pas répondre à "que faut-il pour percer le mur", parce que ces mots n'apparaîtront jamais dans aucune fiche, aussi bien rédigée soit-elle. Les questions posées avec ses propres mots, qui représentaient sur ce catalogue environ un tiers du total, exigent un moteur qui travaille sur le sens.

Les données remises en ordre sont ce qui permet au nouveau moteur de rendre au maximum. Le nouveau moteur est ce qui transforme ce travail en réponses.

---

CASO STUDIO

# Ce que ces chiffres ne disent pas

Il faut le dire clairement, car c'est la limite de la mesure et quiconque sait lire un banc d'essai le demanderait.

Nous avons mesuré que le moteur **répond**, non qu'il répond **bien**. Les quatre premiers résultats des deux moteurs coïncident sur cinq requêtes sur cent vingt et une: nous avons change presque tout le classement, et pour établir que le nouveau classement est meilleur il faut l'oeil de celui qui connaît la marchandise, sur un échantillon.

La donnée solide est celle-ci: auparavant trente-six questions sur cent vingt et une ne recevaient aucune réponse, désormais elles en reçoivent toutes une, et le temps d'attente dans le pire des cas est dix-neuf fois plus bas.

CASO STUDIO

# En résumé

Les trois problèmes trouvés au départ, et le remède à chacun:

| PROBLÈME | EXEMPLE | REMÈDE |
|---|---|---|
| le moteur est ancien | `tournevis` ne trouve pas les tournevis parce qu'il cherche par préfixe | changer de moteur |
| les données sont mal écrites | le titre dit `AEG BBH 12` et ne dit pas que c'est un perforateur | réécrire le catalogue |
| le client décrit son problème | `que faut-il pour percer le mur` | recherche sémantique |

Ce sont trois travaux différents, avec trois coûts différents, et ils peuvent même être abordes l'un après l'autre. Mais ils doivent être faits tous les trois: la reconstruction des données rend trouvable ce qui existe, le nouveau moteur répond à celui qui ne sait pas comment cela s'appelle, et ensemble ils font passer les questions sans réponse de trente-six à zéro.

Le sens comble les lacunes.
**Les mots préservent la précision.**

| 36 → 0 | 19× | 0,3 |
|---|---|---|
| REQUÊTES SANS RÉPONSE | PIRE TEMPS RÉDUIT | RATIO SÉMANTIQUE RETENU |

MI

Massimo Ivaldi

A SEARCH COMPANY

EN RÉSUMÉ

# Trouver davantage.

# Se tromper moins.

La recherche sémantique fonctionne lorsque le catalogue, le modèle et la mesure sont conçus ensemble. Le sens comble les lacunes ; les mots préservent la précision des références et des noms techniques.

01 / Reconstruire

Rédiger le catalogue dans la langue du client, pas seulement dans celle du fabricant.

02 / Mesurer

Utiliser des requêtes réelles et un jeu de test stable, pas des exemples choisis pour réussir.

03 / Équilibrer

Mélanger les mots et le sens dans la proportion adaptée à chaque catalogue.

04 / Vérifier

Contrôler les erreurs, les références, les pires temps et la qualité du classement.

massimoivaldi.com · A search company

septembre 2026