Massimo Ivaldi

A SEARCH COMPANY

GUIDE PRATIQUE

Les 10 règles fondamentales de la recherche interne

Guide pratique, valable pour tout moteur, pour configurer une recherche qui permet vraiment de trouver



massimoivaldi.com

AUTEUR Massimo Ivaldi

Les 10 règles fondamentales de la recherche interne

Synthèse

Une recherche interne fonctionne quand l'utilisateur trouve ce qu'il cherche du premier coup, même en faisant des fautes, avec ses propres mots et sans savoir comment les contenus sont organisés. Le moteur compte moins que sa configuration : presque tous les outils du marché offrent les mêmes leviers de base, et la différence tient à la façon de les régler.

Ce document rassemble dix règles valables pour n'importe quel moteur :

1. Décider ce qui est recherchable, ce qui est filtrable et ce qui est affiché.
2. Tolérer les fautes de frappe et les variantes linguistiques.
3. Traiter la pertinence comme une décision métier, pas comme un réglage par défaut.
4. Proposer des filtres, des facettes et des tris cohérents avec la façon dont l'utilisateur choisit.
5. Répondre immédiatement et montrer pourquoi un résultat est pertinent.
6. Ne jamais laisser l'utilisateur face à une page vide.
7. Renvoyer des résultats propres : sans doublons, regroupés et paginés.
8. Montrer à chacun uniquement ce qu'il a le droit de voir.
9. Mesurer ce que cherchent les utilisateurs et corriger en conséquence.
10. Garder l'index à jour et le système stable.

Introduction : des règles, pas des fonctionnalités

Celui qui utilise le champ de recherche a une intention précise et peu de patience. Sur un site e-commerce, un portail documentaire ou un intranet, la recherche est souvent le chemin le plus court vers une vente, une réponse ou un document. Quand elle échoue, l'utilisateur ne pense pas « le contenu n'existe pas » : il pense « ici on ne trouve rien », et il part.

Les fiches techniques des moteurs de recherche listent des dizaines de fonctionnalités aux noms variés. Sous les noms, pourtant, les concepts sont presque toujours les mêmes : un index, des règles pour comparer la requête aux documents, des règles pour les classer, des outils pour les restreindre. Ce document sépare ce qui est propre à un produit de ce qui est un principe général, et se concentre sur le second.

L'objectif est pratique : donner à quiconque configure une recherche interne, de quelque type que ce soit, une liste de décisions à prendre consciemment plutôt que de les laisser aux valeurs par défaut.

Règle 1 — Définir le rôle de chaque champ

Chaque champ d'un document a un rôle différent, et il doit être déclaré explicitement. Tout indexer « par sécurité » dégrade la pertinence : un mot trouvé dans un code interne ou une note technique fait remonter des résultats que l'utilisateur ne voulait pas.

Pour chaque champ, quatre décisions s'imposent :

Rôle	Question à se poser	Exemple
Recherchable	L'utilisateur taperait-il ce texte dans le champ ?	Titre, description, marque
Filtrable	L'utilisateur veut-il restreindre les résultats sur cette valeur ?	Catégorie, prix, date, langue
Triable	L'utilisateur veut-il trier sur cette valeur ?	Prix, date, popularité
Affiché	Faut-il le montrer dans la page de résultats ?	Titre, image, extrait

Trois précautions valent presque toujours :

- **Peu de champs recherchables, par ordre d'importance.** Le titre pèse plus que la description, la description plus que le corps du texte.
- **Pas de données sensibles ou internes dans l'index** si elles ne servent pas à la recherche : ce qui est indexé peut apparaître dans un résultat.
- **Des données préparées avant l'indexation.** Texte nettoyé du balisage, valeurs normalisées (même format pour les dates, unités, marques), documents longs découpés en sections si l'utilisateur cherche des réponses précises.

Règle 2 — Comprendre l'utilisateur même quand il écrit mal

Les utilisateurs font des fautes, utilisent des synonymes, des abréviations, des singuliers et des pluriels. Une recherche qui exige le mot exact échoue précisément sur les requêtes les plus courantes.

Les leviers de base, présents dans presque tous les moteurs :

- **Tolérance aux fautes de frappe.** Accepter des mots qui diffèrent d'une ou deux lettres. À calibrer : sur les mots courts, une faute change le sens, et sur les références produit ou les nombres, elle est souvent à désactiver.
- **Normalisation.** Majuscules, accents et ponctuation ne doivent pas compter : « eleve », « élève » et « ÉLÈVE » sont la même recherche.
- **Variantes morphologiques.** Ramener les mots à leur racine (« chaussure » et « chaussures », « configurer » et « configuration »). Cela dépend de la langue : il faut définir la bonne pour chaque contenu.
- **Synonymes.** Un dictionnaire tenu à la main, alimenté par les recherches réelles des utilisateurs : « portable » et « smartphone », « congés » et « vacances », sigles et leurs développements.
- **Mots vides.** Articles et prépositions (« le », « de », « pour ») pèsent peu ou rien dans la comparaison, mais sont à manier avec prudence là où ils comptent (« être ou ne pas être »).
- **Recherche par préfixe.** Renvoyer des résultats pendant que l'utilisateur tape encore le mot.
- **Expression exacte.** Les guillemets permettent à l'utilisateur expert de chercher une suite précise de mots.

La règle pratique : large tolérance sur le texte libre, rigueur sur les codes, les nombres et les identifiants.

Règle 3 — La pertinence est un choix, pas un réglage par défaut

L'ordre des résultats est la décision la plus importante de la configuration, car l'utilisateur ne regarde presque que les premiers. Chaque moteur a un classement par défaut fondé sur le texte ; aucun moteur ne connaît les priorités de votre organisation.

Les critères textuels de base sont semblables partout :

- combien de mots de la requête apparaissent dans le document ;
- combien de fautes de frappe ont été nécessaires pour les trouver ;
- à quelle distance ils sont les uns des autres ;
- dans quel champ ils apparaissent (le titre avant le corps) ;
- si la correspondance est exacte ou partielle.

S'y ajoutent les **signaux métier**, à définir par ceux qui connaissent le domaine : popularité, disponibilité, marge, date de publication, autorité de la source. Certains moteurs les combinent en un score, d'autres en une suite ordonnée de critères ; le principe est le même.

Deux outils complémentaires :

- **Résultats épinglés.** Pour quelques requêtes stratégiques, imposer un résultat en tête (« horaires » → page des horaires).
- **Pénalisations.** Rétrograder les contenus obsolètes, épuisés ou archivés sans les supprimer.

Chaque modification doit être vérifiée sur un jeu fixe de requêtes de test avant la mise en production : un ajustement qui améliore une recherche en dégrade souvent d'autres.

Règle 4 — Aider à restreindre, pas seulement à chercher

Beaucoup d'utilisateurs partent d'un mot générique (« veste », « contrat ») puis affinent. Filtres et facettes transforment une longue liste en un choix gérable.

- **Filtres.** Des conditions sur des valeurs structurées : catégorie, fourchette de prix, date, auteur, langue. Ils doivent fonctionner avec la recherche textuelle, pas à sa place.
- **Facettes.** Les valeurs disponibles pour chaque filtre, avec le nombre de résultats à côté (« Rouge (12) »). Elles montrent à l'utilisateur ce qui existe avant qu'il clique, et évitent les combinaisons qui ne donnent aucun résultat.
- **Tri.** Par pertinence par défaut ; par prix, date ou popularité quand l'utilisateur le demande. Trier par date détruit la pertinence : il faut le proposer, pas l'imposer.
- **Filtres géographiques.** Là où la position compte (magasins, sites, services), presque tous les moteurs permettent de filtrer ou trier par distance.

Bonnes pratiques courantes :

- peu de facettes, choisies selon la façon dont les utilisateurs décident vraiment ;
- des valeurs propres et cohérentes (une marque écrite de trois façons devient trois entrées) ;
- des filtres actifs toujours visibles et faciles à retirer.

Règle 5 — Répondre tout de suite et montrer pourquoi

La rapidité fait partie de la pertinence : une réponse lente est perçue comme une mauvaise réponse. Les recherches « pendant la saisie » demandent des temps de réponse de l'ordre de quelques dizaines de millisecondes pour paraître instantanées.

Pour la page de résultats, quelques règles communes :

- **Surlignage.** Les mots trouvés doivent être mis en évidence : l'utilisateur comprend d'un coup d'œil pourquoi un résultat est là.
- **Extrait.** Pour les documents longs, montrer le passage qui contient les termes cherchés, pas les premières lignes.

- **Uniquement les champs utiles.** Ne renvoyer au client que ce qui est affiché : des réponses plus légères, plus rapides, plus sûres.
- **Suggestions en temps réel.** Complétion de la requête et aperçu des premiers résultats pendant la frappe.
- **Cohérence entre appareils.** La même recherche doit donner les mêmes résultats sur ordinateur et mobile ; seule la présentation change.

Règle 6 — Jamais d'impasse

Une page « aucun résultat » est l'endroit où la recherche perd le plus d'utilisateurs. Elle doit être traitée comme un cas à concevoir, pas comme une erreur.

Les stratégies, de la plus simple à la plus élaborée :

1. **Assouplir la requête.** Si aucun document ne contient tous les mots, montrer ceux qui en contiennent le plus, en le signalant à l'utilisateur. Beaucoup de moteurs le font automatiquement, en retirant d'abord les mots les moins importants.
2. **Proposer une correction.** « Vouliez-vous dire… » quand la requête ressemble à une faute de frappe.
3. **Suggérer des alternatives.** Recherches populaires, catégories principales, contenus associés.
4. **Offrir une issue.** Un contact, un formulaire, un lien vers la navigation par catégories.

Chaque recherche sans résultat doit être enregistrée : c'est la liste la plus précieuse pour savoir quels synonymes ajouter et quels contenus manquent.

Règle 7 — Des résultats propres et navigables

Une liste pleine de doublons ou de variantes du même objet donne une impression de pauvreté, même quand la recherche est précise.

- **Dédoublonnage.** Le même produit en dix tailles, ou le même document en trois versions, ne doit apparaître qu'une fois. Presque tous les moteurs permettent de regrouper les résultats selon un champ (référence article, document parent) et d'en renvoyer un par groupe.
- **Variantes dans le résultat.** Tailles, couleurs ou versions s'affichent dans la fiche, pas comme des résultats séparés.
- **Pagination ou défilement.** Le nombre total de résultats doit être indiqué, même approximatif. Au-delà de quelques centaines de résultats, personne ne fait défiler : mieux vaut orienter vers les filtres.

- **Recherche sur plusieurs sources.** Quand les contenus sont dans des index différents (produits, articles, FAQ), on peut les interroger ensemble et les afficher en sections séparées, ou les fusionner en une liste unique. Certains moteurs le font nativement, d'autres demandent de le faire dans l'application.

Règle 8 — Chacun ne voit que ce qu'il a le droit de voir

La recherche est un moyen très rapide de faire fuiter des informations : un titre ou un extrait dans un résultat suffit à révéler un document confidentiel. Dans un intranet ou un service à plusieurs clients, les permissions doivent s'appliquer dans la recherche, pas après.

- **Filtres de sécurité côté serveur.** Chaque requête porte un filtre (service, client, rôle) que l'utilisateur ne peut pas modifier. Filtrer les résultats dans l'interface, après les avoir reçus, ne suffit pas : les décomptes et les facettes révéleraient aussi ce qui n'est pas affiché.
- **Clés à privilèges minimaux.** Clés en lecture seule pour le public, clés d'écriture uniquement pour les systèmes qui indexent, clés d'administration jamais exposées dans le navigateur.
- **Séparation des clients.** Dans un service partagé, des index séparés ou un filtre obligatoire par client. Beaucoup de moteurs proposent des clés ou des jetons qui intègrent le filtre.
- **Permissions à jour.** Quand un utilisateur perd un accès, la recherche doit refléter le changement rapidement.

Règle 9 — Mesurer, corriger, recommencer

Une recherche ne se configure pas une fois pour toutes : elle s'améliore en lisant ce que font les utilisateurs. Les requêtes sont la voix du client, écrite avec ses propres mots.

Indicateur	Ce qu'il mesure	Que faire s'il se dégrade
Recherches sans résultat	Requêtes qui ne trouvent rien	Ajouter synonymes, contenus manquants, corrections
Taux de clic	Part des recherches suivies d'un clic	Revoir classement et présentation
Position moyenne du clic	À quelle profondeur se trouve ce qui sert	Revoir les critères de pertinence

Indicateur	Ce qu'il mesure	Que faire s'il se dégrade
Recherches reformulées	Combien de fois l'utilisateur réessaie aussitôt	Étudier la première requête : que manquait-il ?
Abandon après recherche	Sorties sans clic	Vérifier résultats vides ou hors sujet
Conversion depuis la recherche	Achats, téléchargements, tickets évités	Mesure l'impact métier
Temps de réponse	Latence perçue	Réduire les champs, optimiser index et infrastructure

Une méthode simple et répandue :

1. Chaque mois, extraire les requêtes les plus fréquentes et celles sans résultat.
2. Pour les 50 premières, vérifier à la main si les premiers résultats sont les bons.
3. Corriger synonymes, poids et résultats épinglés.
4. Refaire la vérification sur le même jeu de requêtes pour s'assurer de n'avoir rien dégradé.

Règle 10 — Un index frais et un système stable

Un résultat qui mène à un produit épuisé ou à une page supprimée coûte autant de confiance qu'un mauvais résultat.

- **Mises à jour incrémentales.** N'indexer que ce qui change, dès que cela change, au lieu de tout reconstruire chaque nuit. Prix et disponibilités exigent des délais plus courts que les textes.
- **Suppressions propagées.** Ce qui disparaît de la source doit disparaître de l'index.
- **Reconstructions sans interruption.** Quand il faut tout réindexer (nouveaux réglages, nouveau schéma), construire un nouvel index à côté de l'ancien et les échanger à la fin.
- **Réglages versionnés.** Synonymes, poids et champs sont enregistrés comme configuration, pas modifiés à la main en production : on peut ainsi revenir en arrière.
- **Environnement de test.** Toute modification de pertinence s'essaie sur une copie avant d'être appliquée.
- **Réplicas et sauvegardes.** Des copies pour absorber les pics et une sauvegarde régulière de l'index et des réglages.
- **Supervision.** Latence, erreurs, files d'indexation en retard : des alertes avant que les utilisateurs ne s'en aperçoivent.

Les fonctions de base, en un coup d'œil

Presque tous les moteurs de recherche proposent ces fonctions, parfois sous des noms différents. Le tableau les relie à la règle qu'elles servent.

Fonction	À quoi elle sert	Règle
Champs recherchables et leur poids	Chercher seulement là où cela a du sens, dans le bon ordre	1, 3
Champs affichés	Ne renvoyer que ce qui est montré	1, 5
Tolérance aux fautes	Trouver malgré les fautes de frappe	2
Analyse linguistique (racines, accents)	Traiter les variantes d'un même mot comme un seul	2
Synonymes	Relier des mots différents de même sens	2, 6
Mots vides	Ignorer articles et prépositions	2
Recherche par préfixe	Des résultats pendant la saisie	2, 5
Expression exacte	Chercher une suite précise de mots	2
Règles de classement	Décider qui passe en premier	3
Résultats épinglés	Imposer la réponse aux requêtes clés	3
Filtres	Restreindre selon des valeurs structurées	4, 8
Facettes avec décomptes	Montrer les choix disponibles	4
Tri par champ	Prix, date, popularité	4
Recherche par distance	Des résultats proches d'un point	4
Surlignage et extraits	Montrer pourquoi un résultat est pertinent	5
Regroupement par champ	Éliminer doublons et variantes	7
Recherche multi-index	Interroger plusieurs sources ensemble	7
Clés et jetons avec permissions	Limiter qui voit quoi	8
Statistiques de recherche	Comprendre ce que cherchent les utilisateurs	9
Indexation incrémentale	Garder l'index à jour	10

Certains moteurs ajoutent des fonctions plus avancées, comme la **recherche sémantique** (qui compare le sens plutôt que les mots) ou la **recherche hybride** (qui combine sens et mots-clés). Elles sont utiles quand les utilisateurs décrivent un besoin au lieu de nommer un objet, mais elles ne remplacent pas les règles précédentes : elles se construisent par-dessus.

Checklist et conclusions

Avant de mettre une recherche interne en production, chaque point devrait avoir une réponse réfléchie :

- [ ] Pour chaque champ, il est décidé s'il est recherchable, filtrable, triable, affiché
- [ ] Les champs recherchables sont classés par ordre d'importance
- [ ] La langue des contenus est définie et la tolérance aux fautes est calibrée (désactivée sur les codes et les nombres)
- [ ] Un premier dictionnaire de synonymes existe
- [ ] Les critères de classement incluent les signaux métier
- [ ] Les requêtes stratégiques ont un résultat épinglé
- [ ] Filtres et facettes reflètent la façon dont les utilisateurs choisissent, avec des valeurs propres
- [ ] Les mots trouvés sont surlignés et les extraits sont pertinents
- [ ] La page sans résultat propose des alternatives
- [ ] Les doublons sont regroupés
- [ ] Les permissions sont appliquées côté serveur, avec des clés à privilèges minimaux
- [ ] Les recherches et les clics sont enregistrés
- [ ] Il existe un jeu fixe de requêtes de test
- [ ] L'index se met à jour de façon incrémentale et les réglages sont versionnés

Le moteur choisi compte moins qu'on ne le croit. Presque tous offrent les mêmes leviers ; la qualité de la recherche dépend de ceux qui les règlent, de la propreté des données et de l'habitude de regarder chaque mois ce que cherchent vraiment les utilisateurs.