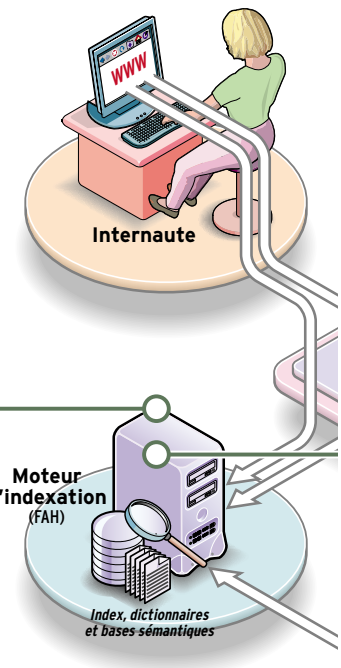
SITES WEB

Moteurs de recherche : l'offre hébergée séduit

Rendus indispensables par la multitude de documents mis sur les sites, les moteurs de recherche sont de plus en plus déployés en mode hébergé. Mais les logiciels traditionnels restent, dans certains cas, incontournables.

Par **Frédéric Bordage**



Tous les outils ne gèrent pas encore XML

▸Mis à part quelques logiciels haut de gamme, parmi lesquels ceux de Verity et de Sinequa, la plupart des moteurs d'indexation proposent une prise en charge limitée et parfois même inexistante de XML. À noter que certains services FAH tels que Synomia prennent en compte l'indexation de documents XML.

SI VOUS ÊTES PRESSÉ

Recourir à un moteur d'indexation améliore le confort des utilisateurs en leur proposant un mode d'accès alternatif aux informations publiées sur les sites web. En plus des architectures client-serveur traditionnelles, les entreprises peuvent désormais exploiter des services d'indexation en mode FAH. Rapide à mettre en œuvre et exploitant un modèle locatif qui répartit les charges liées au moteur de recherche, cette solution séduit un nombre croissant d'entreprises. Elle n'est cependant pas toujours adaptée à des contraintes techniques fortes comme la nécessité d'intégrer des bases de données dans la recherche.

Nous avons 2000 articles en ligne, cela représente donc beaucoup d'informations avec de nombreux points d'entrée. Nous avons fait un effort particulier quant à l'ergonomie de notre site, mais il y a quand même de quoi s'y perdre », avoue Sébastien Leclere, responsable informatique de la Fédération française de motocyclisme. La Fédération a donc décidé de mettre en œuvre un moteur de recherche. Une démarche que toutes les entreprises entreprennent un jour ou l'autre face à la croissance du volume d'informations publiées sur leur site institutionnel ou sur leur boutique en ligne. Le moteur de recherche est alors l'outil idéal pour proposer un mode d'accès alternatif aux informations déjà publiées. Simple à utiliser et ne nécessitant aucun apprentissage, il augmente le confort et fidélise l'internaute. Un moteur de recherche fédère également des ressources hétérogènes et dispersées.

« La multiplication des sites éducatifs et la difficulté d'accès à leur contenu, différentes adresses et différents menus de navigation, nous ont poussés à créer Spinoo, le moteur de recherche du ministère de l'Éducation nationale », explique Erik Boucher, chef de la division du développement numérique au Centre national de documentation pédagogique (CNDP). De son côté, le site marchand Photo12.com possède plus de 180 000 photos numériques en ligne. Traitant de thèmes aussi divers que l'art au XVII[e] siècle ou le terrorisme, chacun de ces clichés peut prendre des significations multiples, ce qui rend impossible un classement dans une seule rubrique. Le moteur de recherche est alors l'outil privilégié pour accéder à un cliché en fonction de ses différentes dimensions. Si l'usage et l'apport des moteurs de recherche sont clairement identifiés par les entreprises, la mauvaise qualité des outils d'indexation livrés avec les serveurs d'applications ou proposés par l'hébergeur les poussent à choisir des solutions commerciales. *« Le petit moteur de recherche Microsoft Index Server, fourni par l'éditeur, ne nous donnait pas satisfaction »*, explique Pierre Petitgas, chargé de communication et webmestre de l'Agence de l'eau Seine-Normandie (AESN). Même constat pour la CFDT. *« Nous ne pouvions pas intervenir sur le moteur de recherche fourni par notre hébergeur. Or, les résultats étaient décevants. Nous avons donc opté pour la solution en mode FAH* [fourniture d'applications hébergées, Ndlr] *Synomia Search »*, explique Véronique Blanc, rédactrice en chef du site cfdt.fr.

L'UTILISATION

Opter pour le mode FAH ou le logiciel serveur

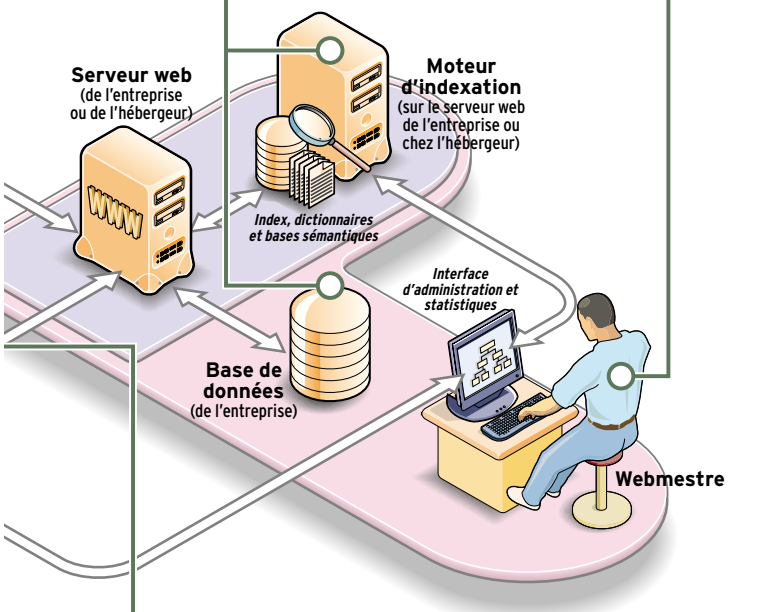
Lorsqu'elles décident de recourir à des moteurs d'indexation professionnels, les entreprises

Des moteurs mieux intégrés

▸Installer un logiciel en local permet d'indexer des bases de données en plus des simples documents HTML, PDF, DOC, PPT, avec d'excellents temps de réponse car le serveur est dédié au site. Il est aussi plus facile d'enrichir des dictionnaires de vocabulaire métier pour obtenir des réponses plus pertinentes. L'intégration, le paramétrage et l'administration peuvent néanmoins être plus lourds qu'un service FAH.

Une interface d'administration de qualité

▸Une interface d'administration de qualité permet de déclencher des indexations, de paramétrer leur fréquence, etc., et fournit des statistiques détaillées sur les requêtes. L'analyse de ces résultats est indispensable pour améliorer la pertinence des recherches, notamment pour les moteurs dotés d'une capacité d'apprentissage.



Un déploiement instantané en FAH

▸Le déploiement d'un service d'indexation en mode FAH ne nécessite aucune intervention technique de la part du site client. Le fonctionnement est identique à celui des moteurs de recherche tels que Google ou Voila. Le moteur d'indexation distant parcourt et indexe régulièrement le site client. L'entreprise n'a qu'à intégrer sa charte graphique à la page de résultats et le tour est joué. Dans la pratique, le service d'indexation prend généralement en charge cette intégration, si bien que le déploiement s'effectue, au pire, en quelques heures.

RETOUR D'EXPÉRIENCE

Niels Stoltenborg

Agence de l'eau Seine-Normandie

▸**Activité :** gestion de l'eau et de l'environnement.
▸**Siège :** Nanterre (92).
▸**Effectif :** 500 personnes.
▸**Budget 2003 :** 850 millions d'euros.
▸**Coût du service Synomia :** 7 500 € ht/an, 625 €/mois.
▸**Trafic web :** 200 recherches par jour.

Pierre Petitgas, chargé de communication et webmestre de l'Agence de l'eau Seine-Normandie.

Choisir la technologie

« Un moteur opérationnel dès le premier contact commercial »

Le site de l'Agence de l'eau Seine-Normandie (AESN) informe le public sur la protection de cette ressource naturelle grâce à 2 000 articles et une base de données sur la qualité des eaux du bassin hydrographique « Seine-Normandie ». *« Le moteur de recherche livré par défaut avec notre site n'était pas satisfaisant et la page d'accueil actuelle ne permet pas au visiteur de se diriger intuitivement vers les rubriques susceptibles de capter son intérêt »,* explique Pierre Petitgas. Pour compenser ces faiblesses, l'AESN exploite depuis quelques mois le service ASP Site Search de Synomia. *« Synomia nous a appelés pour nous présenter son moteur en situation réelle, c'est-à-dire directement sur notre site. Les résultats obtenus étaient pertinents. Le service étant opérationnel, sans installation nécessaire, ni modification chez notre hébergeur, nous avons donné notre accord pour une période d'observation de trois mois. La rédaction des requêtes est simple car en langage naturel. L'analyse linguistique et l'analyse de pertinence permettent d'affiner sa recherche. Les statistiques, essentielles pour comprendre les attentes des internautes, sont aussi un atout important. L'apprentissage en ligne de l'administration du service s'effectue en quelques minutes. Nous avons reconduit notre contrat pour l'année 2004 »,* conclut Pierre Petitgas.

doivent d'abord faire un choix d'architecture, dicté par des aspects techniques et économiques. Si le rôle de ces outils reste fondamentalement le même – construire des index de documents pour faciliter leur recherche – deux offres sont aujourd'hui disponibles : des logiciels installés sur le serveur web de l'entreprise ou de l'hébergeur et des services hébergés par des prestataires externes (mode FAH).

Les outils en mode FAH sont adaptés aux PME possédant des documents relativement homogènes (pages HTML, PDF, etc.), un petit budget et une équipe informatique réduite. *« La possibilité d'utiliser un service externe ne nécessitant aucune installation chez notre hébergeur et pas d'intervention de l'informatique interne, était un critère décisif. Qui plus est, notre moteur de recherche était opérationnel dès l'avant-vente de Synomia. Nous n'avions qu'à signer le contrat pour activer ce nouveau service sur notre site »,* se souvient Pierre Petitgas. Faciles à mettre en œuvre, les services d'indexation proposent des fonctionnalités avancées : langage naturel, analyse linguistique, génération de sous-requêtes pour affiner les résultats, plan de classement dynamique, etc.

Les logiciels spécialisés sont souvent réservés à des besoins plus spécifiques – vocabulaire métier particulier requérant une analyse sémantique poussée, basée sur des dictionnaires spécialisés, très grosse volumétrie de documents à indexer, besoin de marier données **(suite p. 38)**

(suite de la p. 37) structurées (bases de données) et non structurées (HTML, PDF…), etc. Ces solutions haut de gamme exigent alors un budget en licences, en paramétrage et en intégration bien supérieur. *« Avec près de 2 millions de documents hétérogènes répartis sur différents sites et 600 000 requêtes par mois, nous recherchions un logiciel capable de traiter d'importants volumes de documents avec des temps d'indexation et de recherche réduits. C'est pourquoi nous avons retenu K2 de Verity »,* explique Erik Boucher du CNDP. C'est également la rapidité d'Aurweb qui a retenu l'attention de Photo-12.com. Quant au Cridon Nord-Est, avec plus de 400 notaires abonnés accédant à 2 000 documents éminemment techniques, il a préféré le moteur Intuition de Sinequa pour ses capacités d'analyse linguistique, enrichies par des dictionnaires métier.

LA MISE EN ŒUVRE

Cinq minutes pour les moteurs FAH

La richesse de ces logiciels et la pertinence des résultats qu'ils fournissent, nécessitent un travail plus important d'intégration et de paramétrage que les offres FAH. Les éditeurs fournissent alors souvent eux-mêmes les prestations de conseil et d'intégration. *« Nous avons réalisé toute la phase d'installation en étroite collaboration avec Sinequa »,* illustre David Boulanger, directeur du Cridon Nord-Est. L'approche a été la même chez Photo-12.com. Le Centre d'information et de formation des élus locaux (Cidefe) a également fait appel à l'éditeur de son moteur, Auracom, car il ne possédait pas de logistique informatique importante en interne.

Généralement, c'est la création des dictionnaires métier plus que la technique elle-même qui demande le plus de temps. *« Nous disposons d'une base d'équivalences, d'une base d'adjacences d'une base de mots vide ainsi que d'un certain nombre de dictionnaires* [index, Ndlr] *par champ sémantique* [finances, logement…, Ndlr] *qui servent de filtres pour éviter le bruit d'une indexation en texte intégral et la lourdeur d'une indexation manuelle. À chaque fiche sont attribués un ou plusieurs dictionnaires que nous mettons à jour quand de nouveaux concepts apparaissent. Le "Revenu minimum d'activité" a, par exemple, été intégré au Dictionnaire Social »,* explique Claire Riou, ingénieur documentaliste au Cidefe. L'affinage des paramètres systèmes demande parfois une période de rodage. *« L'outil de Verity fonctionne bien "out of the box", mais il faut un très bon niveau technique pour l'exploiter à la hauteur de ses possibilités et de son coût, car c'est un progiciel issu d'un kit de développement. Lors du déploiement, nous avons dû effectuer de nombreux essais pour ajuster tous les paramètres du système »,* confirme Erik Boucher.

Quelle que soit l'architecture retenue, l'intégration avec l'ergonomie du site (charte graphique, navigation, etc.) ne pose aucun problème. *« J'ai envoyé par e-mail à Synomia le formulaire de recherche qui était présent sur notre site et, quelques heures plus tard, l'ensemble était disponible en ligne… Cette mise en place a été d'une rapidité et d'une simplicité déconcertantes »,* constate Pierre Petitgas. Tous les utilisateurs de ces logiciels sont unanimes : le déploiement ne prend que quelques minutes, voire quelques heures dans le pire des cas. *« Pour l'Agence, l'installation du moteur de recherche a consisté à signer un contrat d'utilisation d'un service déjà instantanément disponible en ligne »,* ajoute Pierre Petitgas. Même constat à la CFDT. *« Synomia nous a fourni un lien que nous avons intégré dans notre page d'accueil et quelques instants plus tard, le moteur fonctionnait »,* relève Véronique Blanc.

LES RESSOURCES

Pas de compétence technique requise

L'ajout quotidien de nouvelles pages HTML ou de nouveaux documents est totalement transparent. *« Nous avons créé une macro Word pour générer nos fichiers HTML. Un script les envoie automatiquement en FTP vers le serveur. Entièrement automatisée, l'indexation des documents s'effectue à une fréquence qu'il suffit de paramétrer »,* illustre David Boulanger, du Cridon Nord-Est. Une fois déployés, les moteurs d'indexation ne demandent plus d'intervention technique. Ils s'administrent au travers d'une interface web ou éventuellement par une inter-

RETOUR D'EXPÉRIENCE

Photo12.com/Thierry Foulon

Valérie-Anne Giscard d'Estaing, PDG de Photo12.com.

Photo12
- ►**Activité :** agence photo.
- ►**Siège :** Paris 15e (75).
- ►**Effectif :** 8 personnes.
- ►**Chiffre d'affaires 2003 :** 800 000 €.
- ►**Budget « moteur de recherche » :** 4 000 €/an.
- ►**Trafic web :** 500 clients connectés, 1 500 téléchargements par jour.

Rechercher la performance

« Seul un outil client-serveur offrait la rapidité que nous attendions »

Site marchand de l'agence photo éponyme, Photo12.com propose 180 000 photos numériques en ligne. La société assure également des prestations de gestion de médiathèques numériques pour diverses entreprises, qui regroupent quelque 110 000 photos en ligne. Son site Photo12.com exploite le moteur d'indexation Aurweb de la société Auracom. *« Nous avons retenu cette technologie pour sa rapidité de traitement des requêtes et d'affichage des résultats, et pour sa fiabilité »,* explique Valérie-Anne Giscard d'Estaing, PDG de l'entreprise. Aurweb est en effet éprouvé sur des sites tels que le catalogue de Darty.fr, les sites d'Alapage et de la Fnac, etc. Le moteur fonctionne en full text et indexe la plupart des champs de la base de données SQL Server de Photo12. *« Cela permet des recherches relativement sophistiquées tout en garantissant un temps de réponse exceptionnel »,* constate Valérie-Anne Giscard d'Estaing. Revers de la médaille, Aurweb est assez rigide en termes d'intégration graphique. Mais, comme il constitue la principale fonctionnalité des deux sites, *« nous avons créé notre charte graphique en fonction du moteur »,* explique Valérie-Anne Giscard d'Estaing.

face client-serveur pour certains moteurs d'indexation locaux. *« L'indexation étant automatique, des compétences de documentaliste, ou une certaine connaissance des domaines concernés, sont nécessaires mais uniquement au moment de la création des index et éventuellement pour leur actualisation »*, confirme Claire Riou, du Cidefe. L'interface d'administration fournit également des statistiques qui permettent d'améliorer les recherches en les orientant. *« Nous surveillons le type de requêtes des utilisateurs et le nombre de résultats qu'ils obtiennent »*, illustre Erik Boucher, du CNDP. L'analyse des statistiques offre la possibilité alors de mettre en avant certains contenus ou de guider l'utilisateur lors de sa recherche. *« Synomia Search permet d'insérer des résultats orientés pour envoyer les visiteurs vers certaines rubriques en priorité »*, explique Sébastien Leclere, de la FFMoto.

LES ÉCUEILS

Bien préparer les documents

Le coût des licences des logiciels serveurs peut représenter un frein pour les petites structures. *« Même si les performances sont à la hauteur, le coût de licence de Verity reste élevé et nous avons dû former les développeurs et les administrateurs »*, constate Erik Boucher, du CNDP. Pour mettre en œuvre et administrer Aurweb, Photo12.com a également dû faire appel aux compétences d'un professionnel du langage SQL. Mais finalement, c'est surtout la reprise de l'existant qui peut poser des problèmes. *« La seule difficulté que nous avons rencontrée n'était pas liée au moteur de recherche mais plutôt à la structure de nos pages HTML. Nous avons donc profité de la mise en œuvre de Synomia Search pour créer une charte d'écriture des « balises »* [keywords, titles, etc., Ndlr] *et les avons entièrement passées en revue. Cela nous a permis de faire d'une pierre deux coups en améliorant considérablement notre référencement sur les outils de recherche »*, explique Véronique Blanc.

LES GAINS

Un meilleur service pour l'internaute

Au final, et quelle que soit l'architecture retenue, *« le principal avantage est de pouvoir proposer un service supplémentaire à nos adhérents, accessible en permanence »*, résume David Boulanger du Cridon Nord-Est. L'équipe du site cfdt.fr apprécie également l'économie de temps procurée par un service en mode FAH. *« Nous pouvons ainsi nous concentrer sur d'autres missions, comme la mise en ligne d'informations à partir des requêtes sans résultats des internautes »*, conclut Véronique Blanc. En forte progression, cette architecture en mode FAH devrait connaître un succès croissant car *« il est bien plus facile, rapide et moins risqué d'apprécier, de juger et de mettre en place un service hébergé qu'un service équivalent à installer »*, constate Pierre Petitgas, de l'AESN. ■

AVIS D'INTÉGRATEUR

Thibault Lecuyer, consultant pour Le Projet Web.

« Le mode FAH permet un ROI plus rapide »

Comment choisir entre logiciel et service FAH ?
En ne survalorisant pas l'intérêt du moteur pour le site. Une PME a intérêt à exploiter une offre hébergée pour son site institutionnel. Même si les temps de réponse peuvent être légèrement plus longs, ce service sera moins cher tout en lui apportant plus de souplesse : mise en œuvre rapide, loyer mensuel plutôt qu'investissement, pas besoin de compétences techniques en interne, etc. Au final, les offres de location ont un ROI plus rapide et sans prise de risque technique et financier.

Quand vaut-il mieux privilégier une approche logicielle ?
Un site marchand a, par exemple, tout intérêt à investir dans une offre logicielle car le catalogue est une fonctionnalité centrale pour laquelle aucun compromis ne doit être fait. Les temps de réponse doivent être les plus courts possible et les contraintes techniques inhérentes à ce type de projet, base de données sous-jacente, trafic web, etc., sont peu adaptées à une approche de type hébergée.

Le Projet Web
- **Activité :** société de conseil en stratégie web et de conception de sites.
- **Siège :** Tours (37).
- **Effectif :** 2 personnes.
- **Références :** Nestlé Waters, ministère de la Culture, IGN, OFPRA.

Les principaux moteurs d'indexation pour sites web

Éditeur/Site web	Logiciel	FAH/logiciel	Fonctionnalités clés	Prix (à partir de)
Albert www.albert.com	AMI Website Access	FAH/logiciel	Recherche en langage courant sur données structurées et non structurées. Tous formats de fichiers, dont XML. Apprentissage automatique. Détection automatique des concepts-clés. Documents similaires. Fédération de contenus multisources. Mécanisme de « tête de gondole » pour site d'e-commerce.	Licence : 30 000 € ht FAH : 3 000 € ht
Sinequa www.sinequa.com	Intuition	FAH/logiciel	Moteur de recherche et de navigation sémantique multilingue (plus de dix langues acceptées). Combine les technologies texte intégral booléen, statistique, recherche structurée multicritère et linguistique. Gestion native XML et intégration des ressources documentaires existantes (plans de classement, taxinomies...).	Serveur : 30 000 € ht FAH sur devis
Synomia www.synomia.fr	Synomia Search	FAH	Recherche multicritère sur données structurées et non structurées. Tous formats de fichiers, dont XML. Plan de classement dynamique. Analyse linguistique. Résultats orientés. Déploiement en 5 minutes. Statistiques.	190 € ht/mois
Auracom www.auracom.fr	Aurweb	Logiciel	Recherche multicritère sur données structurées et non structurées. Plan de classement dynamique. Dictionnaire métier pour affiner les recherches.	2 800 € ht/an
Verity www.verity.fr	Verity K2	Logiciel	Fédération de sources internes et externes. Recherche, classification, catégorisation, moteur de recommandation (documents, produits, utilisateurs, etc.). Multilinguisme : 90 langues acceptées. Respect de la sécurité.	30 000 €/ht pour 2 CPU pour 2 ans
	Verity Ultraseek	Logiciel	Moteur de recherche. Installation faite en moins de 2 minutes. Administration simple. Interface personnalisable.	7 500 € ht pour 5 000 documents