

"Site Search" : les principales offres (2ème partie)[Retour au sommaire de la lettre](#)

Pour indexer les contenus internes à un site Web, vous avez le choix entre une solution de recherche en mode ASP, une application sur serveur ou un boîtier de recherche intégré. Différentes solutions d'analyse des résultats sont également proposées. Quelle solution choisir ? Nombreux sont les webmasters qui se sont posés ou se posent actuellement cette question. Chaque type de solution a bien entendu des avantages et des inconvénients. Petit tour d'horizon des offres des principaux acteurs. Ce panorama vient compléter celui, plus général, qui était consacré le mois dernier au marché de la recherche intra-site.

Les fournisseurs de solutions de recherche sont si nombreux qu'il est parfois extrêmement difficile de s'y retrouver (voir la première partie de cet article dans la Lettre R&R de mars 2006). Faut-il opter pour une solution en mode hébergé (ou ASP), pour une application à installer sur serveur ou plutôt adopter un boîtier "tout-en-un" de type "Google Appliance" ? Quel que soit le type de solution retenue, une attention particulière doit être portée aux fonctionnalités de recherche, et notamment aux modes de recherche syntaxiques, sémantiques ou statistiques.

"Le choix est toujours délicat, explique Laurent Baleyrier, co-fondateur de la société Kartoo (<http://www.kartoo.com/>). Tout dépend de la volumétrie du site et de la façon dont l'information est organisée. Il faut également réfléchir à la manière la plus adaptée pour organiser et qualifier l'information". Les moyens financiers et humains à votre disposition sont un autre paramètre à prendre en compte, bien entendu.

Solutions serveurs ou solutions en ASP ?**Caractéristiques des principaux moteurs**

Nous avons volontairement écarté ici les solutions open source et celles qui sont intégrées à des gestionnaires de contenus en open source comme SPIP (<http://www.spip.net/>). Pour toute information sur les solutions libres de recherche, voir la Lettre R&R de juillet 2005. Voici un tableau regroupant les principaux éditeurs de solutions :

Mode	Editeur	Nationalité	Exs de clients / utilisateurs
ASP	Atomz (http://www.atomz.com/)	US	Lego, palmOne...
	Synomia (http://www.synomia.fr/)	FR	Libération, Renault...
	Verticrawl (http://www.verticrawl.com/)	FR	123-emploi, FraGGo.com...
Licence	Exalead (http://corporate.exalead.com/)	FR	Groupe Afnor, Danone...
	Go Albert (http://www.albert.com/)	FR	Unesco, BNP Paribas...
	Sinequa (http://www.sinequa.com/)	FR	Le Monde, Cdiscount...
	Spirit (http://www.spiritengine.com/)	FR	MMA, Cour des Comptes...
Mixte (licence ou ASP)	Verity (http://www.autonomy.com/)	UK/US	Les Echos...
	Antidot (http://www.antidot.net/)	FR	Journal du Net, TF1...
Appliance	Fast (http://www.fastsearch.com/)	Norvège	Priceminister, Strategies.fr
	Google (http://www.google.fr/)	US	Banque mondiale, Nokia...
Open source	Lucene (http://lucene.apache.org/)	US	Bibliothèque de Rouen, Keljob.com...

Mode	Editeur	Nationalité	Exs de clients / utilisateurs
	Ht://Dig (http://www.htdig.org/)	US	Oxford University, Yale University Library...
	ASPseek (http://www.aspseek.org/)	US	Jabber, Mississippi State University...

Applications hébergées chez le client

Les solutions sous licence sont de loin les plus répandues. Hébergées sur un serveur chez le client, elles sont plus chères que les solutions en mode ASP. Elles sont le plus souvent mises en avant par les fournisseurs pour la possibilité de personnalisation et de contrôle offerte au client et pour leur interopérabilité avec les différents silos de données présents dans l'entreprise.

Dans cette catégorie, le Français Sinequa revendique une place de leader dans plusieurs secteurs, avec sa technologie Intuition. Il est très présent dans la presse, l'e-administration et le secteur du e-commerce. Les solutions des Français Exalead et Antidot sont également très répandues, de même que celles de Verity (racheté par le Britannique Autonomy l'an dernier) et du norvégien Fast, qui a remporté d'importants contrats en France depuis son implantation fin 2005.

Deux questions à Sébastien Joncheray (Raynette – www.raynette.com)

Créée en 1998 et co-développeur de Mozbot (avec Abondance), la société nantaise Raynette développe et commercialise des applications serveurs pour sites Internet de PME (solutions de newsletter, de recherche intra-site, et de e-commerce). Ses revenus proviennent des ventes de licences d'applications-serveurs.

L'entreprise met l'accent sur la simplicité de déploiement de sa solution pour PME, avec une installation sur site. Ses références clients comprennent des TPE (Bio Conseil...), le Ministère de la Défense et des mairies.



Quels sont, selon vous, les critères qu'il faut prendre en compte pour la sélection d'une solution de recherche intra-site ?

Il faut garder une adéquation entre le volume, la complexité du site, et la "dimension" de la solution de recherche. Il y a une tendance à opter pour des outils de recherche surdimensionnés pour les besoins réels. A moyen terme, cela devient pénalisant car cette solution est trop lourde et compliquée à gérer, et, pour finir, la solution est mal utilisée.

Quels seront, d'après vous, les grands enjeux à relever pour les solutions de recherche intra-sites dans les années à venir ?

Il faudra répondre aux nouveaux besoins réguliers d'affinement des recherches et d'expansion du nombre et des types de documents, tout en gardant des solutions maniables qui ne soient pas des "usines à gaz". Nous visons toujours la simplicité pour l'utilisateur de nos produits, et c'est parfois compliqué :-)

Solutions en mode ASP

Stimulées par l'essor du haut débit, les applications de recherche hébergées ont le vent en poupe. Elles sont principalement appréciées pour leur simplicité de déploiement et leur mode locatif, qui ne nécessite pas d'investissement initial dans des licences.

L'administration s'effectue également à l'aide d'un simple navigateur.

Les appliances "clefs en mains"

L'appliance de Google est probablement celle qui fait couler le plus d'encre à l'heure actuelle, du fait de son prix jugé "agressif". Ce boîtier offre un système packagé, englobant le matériel et le logiciel, prêt à l'emploi. Deux modèles sont disponibles, se différenciant par leur capacité de traitement des documents : le Google Mini (pour les petites entreprises, à partir de 50.000 euros pour 2000 documents) et le Google Search Appliance (pour les moyennes et grandes entreprises, jusqu'à 15 millions de documents, pour environ 30 000 euros).

Si les produits de Google sont légèrement critiqués par Gartner, qui considère qu'ils offrent un moindre degré de sécurité et de personnalisation que ceux de ses concurrents, le quotidien L'Expansion estime lui que le groupe remporte actuellement "entre 10 et 20% des nouveaux contrats" (<http://www.lexpansion.com/art/32.0.142256.0.html>).

Google a plusieurs cordes à son arc pour s'attaquer à ce marché. A l'instar d'autres éditeurs du monde du stockage et de la sécurité, il s'efforce d'offrir une brique de recherche "de base" sur laquelle les développeurs tiers sont invités à développer des modules, reproduisant ainsi le modèle qui a fait le succès de Google Maps.

Dernièrement, Google s'est par ailleurs associé à plusieurs éditeurs - dont Cisco, Oracle, Cognos, Salesforce.com et SAS - pour assurer la compatibilité de sa solution avec leurs applications.

Une autre appliance de recherche est par ailleurs proposée par l'américain Thunderstone (<http://www.thunderstone.com/>). Elle est dénommée Webinator.

Connecteur

Et pourquoi ne pas adopter une appliance Google intégrée pour la personnaliser ensuite avec des connecteurs développés par les éditeurs de votre choix ?

Actuellement, la société KartOO développe des connecteurs vers Google Mini, Verity (Autonomy) et d'autres solutions devraient suivre. Objectif : "développer des connecteurs vers des solutions qui existent déjà pour apporter une nouvelle brique technologique supplémentaire sur ces moteurs", nous affirme Laurent Baleyrier, directeur de Kartoo.

Principales technologies de recherche

Que la solution soit proposée en mode serveur ou ASP, on distingue quatre principales technologies :

- La **recherche lexicale** : proposée par tous les moteurs, elle permet de rechercher un mot ou un groupe de mots, avec la possibilité d'effectuer des requêtes booléennes (avec les opérateurs ET, OU, SAUF...) ou des opérateurs d'adjacence et de proximité.
- La **recherche statistique** : très répandue là aussi, cette technique permet de calculer la pertinence de ces mots par rapport aux termes contenus dans l'index (la base) et aux documents répondants à la requête. Les résultats sont ensuite triés par ordre de pertinence.
- La **recherche syntaxique** : ce mode de recherche permet de retrouver un terme quelle que soit sa déclinaison (plurielle ...) ou d'étendre la recherche aux synonymes ou aux mots ayant la même racine étymologique. S'ajoutent à cela le plus souvent des modules de recherche floue ou de phonétisation.
- La **recherche sémantique** : une analyse du sens de la requête est ici effectuée, par le biais d'une recherche des mots sémantiquement proches de ceux qui sont utilisés. La recherche s'appuie alors sur un dictionnaire sémantique qui, à chaque mot de la langue, associe leurs différents sens. Avantage : ce principe offre un grand niveau de souplesse quant au choix des mots employés dans les requêtes. Inconvénient : il nécessite le plus souvent un travail de la part de l'entreprise pour enrichir les dictionnaires.

Les moteurs statistiques et les moteurs sémantiques constituent le gros du marché.

Tableau comparatif des principales méthodes d'analyse et de recherche

Editeur (solution)	Nature des moteurs	
	Statistique	Sémantique
Antidot (Antidot Finder Suite)	X	
Atomz (Atomz Search)		X
Exalead (exalead one:enterprise)	X	
Fast (Enterprise Search Platform)		X
Go Albert (Amy Enterprise Discovery)		X
Sinequa (Intuition)		X
Spirit		X
TripleHop (Matchpoint)	X	
Verity (Ultraseek)		X
Verticrawl (Intra-Site)		X

Sources : d'après Journal du Net (05/10/2005 - http://solutions.journaldunet.com/0510/051005_panorama_moteurs.shtml) ; Abondance

S'ajoutent à ces types de recherche, la recherche structurée XML, la recherche dans des bases de données relationnelles – entre autres sur les sites de e-commerce - et la recherche mathématique (avec une modélisation vectorielle du sens des documents, différente du comptage des co-occurrences de mots de la recherche statistique).

Par ailleurs, si tous les acteurs proposent la recherche plein texte à leurs utilisateurs, quelques moteurs ne proposent pas de recherche phonétique ou de recherche approximative (couplée à un correcteur d'orthographe).

Editeur (solution)	Approximative	Phonétique	Langage naturel	Plein texte
Antidot (Antidot Finder Suite)	X	X		X
Atomz (Atomz Search)	X		X	X
Exalead (exalead one:enterprise)	X	X	X	X
Fast (Enterprise Search Platform)	X	X	X	X
Go Albert (Amy Enterprise Discovery)	X	X	X	X
Sinequa (Intuition)	X	X	X	X
Verity (Ultraseek)	X	X	X	X
Verticrawl (Intra-Site)	X	X	X	X

Sources : d'après Journal du Net (05/10/2005 - http://solutions.journaldunet.com/0510/051005_panorama_moteurs.shtml) ; Abondance

Toutes ces fonctionnalités sont souvent conjuguées pour offrir des modes "avancés" de recherche et de restitution des résultats (résumé de documents, classement avancés et sélection optimisés des réponses...).

Conclusion

L'appliance de Google offre un exemple d'une tendance de fond dans le monde de la recherche intra-site, avec une évolution vers des plateformes simples, aussi larges que possibles, sur lesquelles se grefferont de plus en plus de solutions spécialisées à l'avenir. Les divers rachats intervenus en 2005 ont en outre signalé le début d'un mouvement de consolidation jouant en faveur des "grands" éditeurs, à l'origine d'autres solutions de gestion de contenus : TripleHop (solution MatchPoint) est notamment entré dans le giron du groupe Oracle en juin 2005, Autonomy et iPhrase sont ensuite respectivement tombés dans le giron de Verity et IBM en novembre de cette même année.

"Ce n'est pas qu'un moteur de recherche pour sites Web ou intranet. Nous préférons parler de plate-forme d'accès, de recherche et de navigation au sein de toutes les informations de

l'entreprise", soulignait ainsi dernièrement Jean Ferré, PDG de Sinequa, à propos de la dernière mouture de sa solution Intuition (voir <http://www.01net.com/article/300499.html>), marquant sa volonté d'offrir plus qu'un moteur de recherche. Du côté de Synomia, outre la recherche, l'accent est mis sur les solutions d'interaction client ou visiteur (voir entretien à la suite de cet article).

En conclusion, si la qualité des outils et des fonctionnalités de recherche reste un élément différenciateur important, ne faut-il pas de plus en plus prêter attention à d'autres éléments associés comme les offres de navigation ou de maillage de l'information ou encore à la compatibilité de ces solutions de recherche avec les autres solutions en place dans les entreprises... ?