

"Net Recherche", un livre sur l'art de la recherche d'information sur le Web

[Retour au sommaire de la lettre](#)

L'ouvrage "Net Recherche", paru récemment chez ADBS Editions, nous explique l'art si difficile parfois de la recherche d'informations sur le web. En six chapitres et de nombreuses fiches pratiques, il nous aide à mieux formuler nos requêtes et à gagner du temps dans nos investigations sur la Toile. Voici quelques extraits de ce livre qui vous aideront à vous faire une meilleure idée de son contenu...

Qu'elle réponde à des besoins professionnels ou personnels, la recherche d'information représente une grande part de l'activité des internautes. Pourtant, le fonctionnement des moteurs de recherche est encore mal connu, et les utilisateurs sont souvent perdus face à des résultats trop nombreux et imprécis, à des contenus structurés ou non, validés ou non, gratuits ou payants... L'art de la recherche d'information consiste pour une grande part à déterminer les outils, généralistes ou spécialisés, les plus adaptés aux différents types de requêtes et à les utiliser avec maîtrise.

L'ouvrage "Net Recherche"

(<http://www.adbs.fr/site/publications/ouvrages/104.php>)

a pour objectif d'offrir à toute personne amenée à effectuer des recherches sur Internet un panorama des outils et méthodes existants à ce jour. Il est destiné tant aux professionnels de l'information (documentalistes, bibliothécaires, veilleurs) qu'aux enseignants, chercheurs, étudiants, confrontés à la complexité sans cesse grandissante des données disponibles en ligne. Ils y trouveront des conseils méthodologiques ainsi que des "trucs et astuces" destinés à optimiser le processus de recherche, des informations précises sur les évolutions actuelles ainsi que des présentations d'outils et des listes d'adresses utiles. Le livre est écrit par Armelle Thomas, consultante en veille documentaire et stratégique et Véronique Mesguich, responsable des activités Entreprises de l'Infothèque du Pôle universitaire Léonard de Vinci.

En accord avec l'éditeur, voici quelques extraits de cet ouvrage :



Extrait 1 : Onze tendances actuelles pour les moteurs de recherche (Chapitre 2)

1. Simplification de l'interface et aide à l'utilisateur

Au cours des dix dernières années, le mode de fonctionnement général des moteurs a somme toute peu évolué. En revanche, on constate une tendance à la simplification de l'interface d'interrogation. L'opérateur OU par défaut a laissé la place, sur quasiment tous les moteurs désormais, à l'opérateur ET.

Les options de recherche avancée se sont également allégées, en supprimant au passage quelques fonctionnalités intéressantes comme la troncature. La complexité de la recherche d'information réside désormais davantage dans la maîtrise des contenus que dans une expertise technique des langages d'interrogation. Il faut souligner également un progrès important en ce qui concerne les services à l'utilisateur : reformulation des questions, correcteurs orthographiques, suggestion de mots clés ou de pages similaires, etc., avec des interfaces évolutives.

Des options de présentation des résultats originales (cartographie, prévisualisation) ont également vu le jour.

2. Les raccourcis de recherche

De plus en plus, les moteurs proposent à leurs usagers des raccourcis de recherche pour obtenir directement les informations les plus variées : par exemple en inscrivant "meteo Nantes" ou

"horoscope Lyon" sur Yahoo! France, on obtient directement les informations souhaitées en tête de page de résultats. Les raccourcis ne fonctionnent pas forcément avec un mot clé : ainsi, toujours sur Yahoo!, choisir le nom d'un artiste comme mot clé de recherche permettra d'obtenir immédiatement ses clips ; sur Google, l'insertion d'un nom de lieu dans le moteur ramènera le plan du dit lieu, une recherche avec deux noms de ville ramènera les horaires de train, etc. (se reporter aux modes d'emploi des moteurs pour une liste complète des fonctions).

3. La personnalisation, la proximité avec l'internaute

C'est l'un des grands chantiers pour les outils de recherche (qui leur permet aussi de mieux fidéliser leurs "clients"), et un challenge pour l'avenir. Il s'agit de permettre aux internautes d'interagir avec leur moteur, en leur permettant, au-delà d'une simple personnalisation de l'interface et des préférences (devenue assez classique), de stocker des éléments d'information dans un espace dédié du serveur de l'outil, de conserver un historique de ses recherches (voir page 93), de surveiller des requêtes, de partager de l'information avec d'autres personnes selon le principe de la navigation sociale (voir page 59), etc. En voici quelques exemples.

- Ujiko (<http://www.ujiko.com/>), lancé par l'équipe du métamoteur Kartoo à partir de la technologie Yahoo!. L'outil joue à fond la personnalisation, en permettant la mémorisation et personnalisation des recherches. De plus, les url des documents peuvent être cochés, annotés, filtrés, classés ou supprimés. Lorsqu'on clique sur un des résultats d'une recherche, la page est conservée en mémoire et ultérieurement placée en tête de résultats.

- Ask Jeeves propose My AskJeeves (<http://myjeeves.ask.com/>) qui permet de sauvegarder des liens obtenus comme résultats lors de requêtes sur le moteur et de les gérer par la suite (mention "save" à côté d'un résultat) dans des catégories, comme un bookmark, et de les annoter.

- Yahoo! a lancé My Yahoo! ("Mon Yahoo!" en version française et opérationnelle), qui permet de sauvegarder les résultats, voire de les commenter et d'y effectuer des recherches, ainsi que ses favoris. On peut aussi exclure un lien pour de futures requêtes. La version V2 de l'outil accentue le type de contenus gérables (pages web, fils RSS) et l'aspect partage.

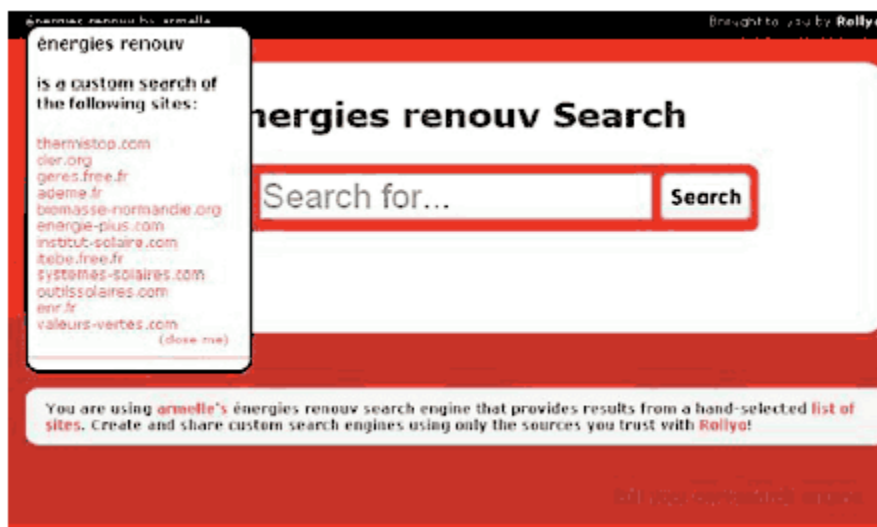
- Yahoo! propose également à l'utilisateur, depuis 2005 et via le site Yahoo! Mindset, d'orienter les résultats de la recherche vers des sites informatifs ou commerciaux, grâce à une sorte de curseur. Ainsi, en fonction de l'action de l'utilisateur, les résultats seront classés selon leur teneur, soit informationnelle, soit commerciale. On peut tester ce nouveau service en version bêta (<http://mindset.research.yahoo.com/>).

- Google a bien entendu présenté lui aussi sa solution Personalized Search qui complète l'outil de sauvegarde des requêtes "My search History". L'outil Google Trends, apparu fin 2005, complète ces options en proposant des statistiques sur les requêtes les plus fréquentes et les sites les plus visités par l'internaute. Ces trois outils nécessitent d'ouvrir un compte de messagerie Gmail (rappelons que Google a racheté en 2003 la société Kaltix, spécialisée dans la personnalisation des recherches). À tester sur <http://labs.google.com/personalized>.

- Mozbot propose diverses options de personnalisation, dont la possibilité d'afficher les "favicons". Le favicon est un pictogramme destiné aux favoris enregistrés dans le navigateur.

- Amazon a lancé le moteur A9 (<http://www.a9.com/>) dont les résultats web sont fournis par Google. Cet outil, également axé sur le catalogue d'ouvrages de la librairie en ligne, permet notamment de conserver un historique de ses précédentes recherches.

- Enfin, l'outil Rollyo, lancé fin 2005, permet de mener des recherches, via le moteur Yahoo!, dans un searchroll, sorte de bouquet de sites prédéfini par l'utilisateur. La création des searchrolls est très simple ; et peut se faire à partir des bookmarks pour les utilisateurs de Firefox. De plus, l'outil permet de s'inspirer d'autres searchrolls créés par divers internautes. On retrouve ici l'illustration d'un des principes majeurs d'Internet : le partage d'informations... (<http://www.rollyo.com/>)



Exemple de moteur spécialisé conçu avec Rollyo

4. Le clustering ou catégorisation automatique

Le terme cluster signifie "regroupement" ou "grappe". Ce terme peut avoir plusieurs significations en informatique. En l'occurrence, cela désigne des techniques statistiques (méthode des mots associés avec matrices de cooccurrences) qui permettent d'extraire automatiquement des termes dans les pages de résultats trouvées par un moteur ou métamoteur, afin d'affiner ou de réorienter une recherche. Sur le web, où l'on ne peut pas faire référence à des dictionnaires ou ontologies préexistantes (type thésaurus ou autre), cette catégorisation effectuée dynamiquement, en fonction de la requête, pourra permettre de lever des ambiguïtés ou de préciser une recherche. Les clusters ainsi formés servent à mettre en évidence des thématiques présentes dans les documents trouvés par les outils de recherche. Cela peut présenter un intérêt pour la recherche documentaire classique (classement automatique des résultats selon des thématiques prédéfinies ou générées automatiquement par le moteur) ou pour la veille (mise en évidence de co-occurrences de termes, ce qui peut permettre ensuite d'en tirer des déductions à caractère stratégique).

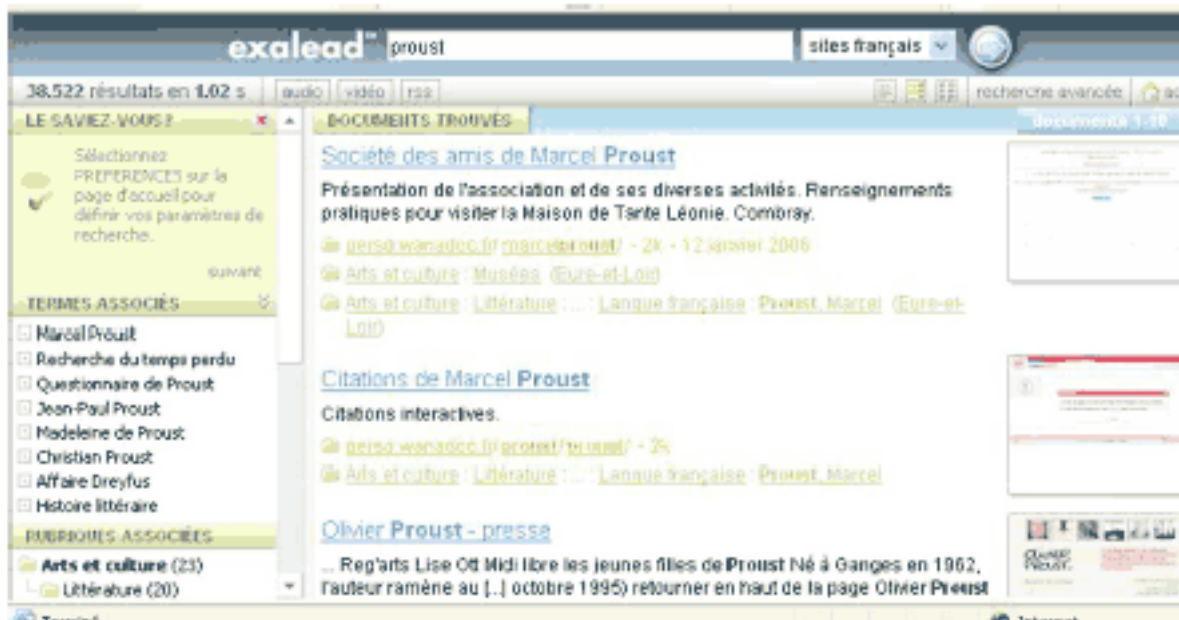
Des moteurs comme Exalead ou Teoma, des métamoteurs comme Vivisimo, Polymeta ou Kartoo utilisent la technologie des clusters, mais l'application peut varier d'un outil à l'autre. Exalead utilise la technologie des clusters pour à la fois générer les "termes associés" qui vont permettre d'affiner la recherche, mais aussi offrir la possibilité de l'affiner à partir d'une liste de "catégories" définie a priori.

Ces termes associés sont générés automatiquement, en fonction de leur fréquence d'occurrence dans le texte intégral des pages trouvées par le moteur. Par contre, contrairement à Teoma ou Kartoo, Exalead ne propose que des expressions en tant que "termes associés", et non pas des mots clés uniques (unitermes, en langage documentaire).

Le moteur Teoma utilise également la technologie des clusters, mais de façon un peu différente. À partir d'un mot clé, Teoma recherche comme tout moteur les pages figurant dans son index et contenant au moins une fois le mot clé. Puis, Teoma classe ces pages dans des ensembles grâce à l'analyse des liens (regroupements des pages pointant les unes sur les autres). Les intitulés de ces ensembles correspondent aux termes les plus communs à ces pages : ce qui donne parfois, là encore, des intitulés un peu curieux, surtout si l'on utilise des termes en français...

Le clustering permettra ainsi de démarrer une recherche à partir d'un terme assez généraliste, et d'avoir une idée de "l'environnement sémantique" correspondant à une recherche, c'est-à-dire les expressions apparaissant le plus fréquemment dans les pages contenant le ou les mots clés de départ. Cela permet aussi de limiter l'ambiguïté sur une recherche : par exemple, on pourrait éliminer d'une recherche sur Proust tout ce qui concerne le préfet Jean-Paul Proust qui n'a rien à voir avec Marcel.

Ces outils de clusterisation peuvent fonctionner comme une brique associée à un autre moteur : exemple avec le logiciel récemment sorti TopGist qui permet de "thématiser" des recherches effectuées avec Google ou Yahoo!



La cartographie selon Exalead

Enfin, Google teste depuis début 2006 un système de classement des résultats sous forme de dossiers thématiques, en fonction de la nature de ces résultats (actualité, blogs, forums, etc.).

5. La cartographie

La représentation de résultats de recherche sous forme de carte visuelle, par rapport à une présentation classique sous forme de liste, vise à apporter une vision d'ensemble plus claire, à mettre en évidence des liens entre pages et mots clés, à appréhender plus aisément des volumes importants de données...

La visualisation est aujourd'hui en progrès, avec des composantes dynamiques et contextuelles. En général, un ensemble d'information n'est pas représenté par une carte unique qui aurait souvent du mal à rendre compte de la complexité de l'ensemble, mais par une multitude de cartes ou vues reliées entre elles.

Les technologies de visualisation s'intéressent à des types de données de plus en plus diversifiés, et à des volumes de plus en plus grands. Elles s'interfacent naturellement avec des moteurs de recherche ou d'autres applications d'analyse de données, notamment les moteurs utilisant des technologies de classification automatique. Si la lisibilité et l'interprétation restent souvent délicates, ces techniques sont néanmoins certainement promises à un bel avenir sur le web (et ailleurs) pour aider à la prise de connaissance d'un corpus, et générer de nouvelles hypothèses de travail.

Exemples :

- technologie Kartoo appliquée à son métamoteur : <http://www.kartoo.com/> ;
- technologie The Brain appliquée aux résultats de l'Open Directory : <http://www.webbrain.com/> ;
- technologie Anacubis appliquée à Google : <http://www.i2.co.uk/anacubis> ;
- technologie Mapstan (Social Computing) appliquée à Societe.com (option de cartographie des sociétés qui possèdent des dirigeants communs) ;
- exemple de navigation graphique sur Renardus (<http://www.renardus.org/>) : choisir un grand thème, un sujet, et cliquer sur "graphical navigation".

6. Les moteurs de recherche personnels (desktop search)

Actuellement, la plupart des grands moteurs (voire métamoteurs, comme Copernic) proposent gratuitement aux internautes de disposer de leur technologie pour indexer le contenu de leur disque dur, voire celui des réseaux internes de l'entreprise, et effectuer des recherches. De nombreux formats de documents peuvent être reconnus.

Voici plusieurs exemples récents :

- Copernic Desktop Search : <http://www.copernic.com/en/products/desktopsearch/index.html> ;
- Google Desktop : <http://desktop.google.com/> ;
- Exalead Desktop : <http://www.exalead.fr/> (version d'évaluation de trente jours).

En septembre 2005, le site américain ZDNet a publié un comparatif de cinq outils desktop (Google, Yahoo, DTsearch, Copernic, MSN Toolbar). On peut consulter les résultats de ce banc d'essai à l'adresse suivante : http://reviewszdnet.com.com/4520-3684_16-6313062.html

La version 3 de Google Desktop, lancée en février 2006, a suscité des réserves de la part de défenseurs de la vie privée, notamment l'Electronic Frontier Foundation. En effet, une option de cette nouvelle version du moteur personnel, dans l'objectif de faciliter les recherches simultanées sur plusieurs ordinateurs, implique d'effectuer des copies temporaires des documents concernés sur un serveur chez Google. Cela pourrait constituer une violation de la vie privée des utilisateurs...

7. Les barres d'outils

Aujourd'hui, la plupart des grands moteurs proposent en téléchargement leur barre d'outils (toolbar) qui s'installe sur le navigateur (malheureusement souvent sur Internet Explorer seulement, mais de plus en plus aussi sur Firefox).

Ces barres offrent alors un certain nombre de fonctionnalités très pratiques, dont la première reste bien sûr la recherche directe sur le web, sans avoir à aller sur le site de son outil préféré. Certaines permettent de rajouter les moteurs de son choix (exemple, celle de Copernic).

Fonctionnalités proposées :

- recherche sur le web (pages, documents multimédias, etc.) ;
- ajout de moteurs de son choix (exemple Copernic toolbar) ;
- recherche au sein de la page visitée ;
- informations sur la page visitée ;
- traduction ;
- mise en surbrillance des termes de la requête ;
- blocage de fenêtres pop-up (exemple Google) ;
- historique de recherche ;
- personnalisation des affichages (exemple : <http://www.toolbarbrowser.com/>).

Développée uniquement pour Mozilla/Firefox, la barre d'outils Needlesearch permet d'interroger tout moteur de recherche généraliste choisi par l'utilisateur, ou encore toute page web disposant d'un moteur de recherche (par exemple, Legifrance). On peut télécharger gratuitement cet outil (www.needlesearch.com).

8. Le modèle économique

L'ère du "tout gratuit" est désormais terminée pour les moteurs de recherche. On constate la pérennisation d'un modèle économique basé sur la publicité. Cela peut prendre plusieurs formes : la diffusion de bandeaux publicitaires sur les sites portails généralistes, ou bien le développement du référencement payant, et surtout du positionnement payant, qui garantit une bonne place dans la liste des réponses (le système Adwords de Google par exemple). Ce principe des liens commerciaux s'est beaucoup développé, pour autant les moteurs sont tenus de distinguer les réponses "classiques" des réponses "sponsorisées".

9. La recherche locale ou géolocalisation

Quelques moteurs permettent d'effectuer des recherches à caractère local, en précisant une recherche par mot clé avec l'intitulé d'une région, d'un département ou d'une ville. C'est le cas de Mirago par exemple (<http://www.mirago.fr/>).

Google, Yahoo et MSN se positionnent également très fortement sur ce marché.

La recherche locale n'est pas aisée sur des pages web classiques, dans la mesure où les métadonnées ne comportent pas de champ qui permettrait une localisation précise. L'avenir serait aux outils de recherche locale basés sur des annuaires de type "pages jaunes", qui pourraient générer des revenus publicitaires importants provenant de PME locales...

10. Les moteurs spécialisés

Ils sont encore peu nombreux, car ils font rapidement appel à des technologies complexes.

Certains font une indexation en texte intégral des pages d'une sélection manuelle de sites web, tels LawCrawler du site Findlaw dans le domaine juridique (<http://lawcrawler.findlaw.com/>) ou Netsearcher (<http://www.netsearcher.com/>) sur une sélection de sites internet.com.

D'autres catégorisent automatiquement des pages web, tel Voila avec sa recherche thématique (à tester par exemple avec le mot "bilan" dans la thématique "comptabilité").

Parmi les moteurs spécialisés, on peut citer Scirus (<http://www.scirus.com/>) : ce moteur, basé sur la technologie Fast Search, indexe en profondeur des sites web d'accès libre et des bases de données gratuites. Seuls des sites à contenu scientifique validé sont sélectionnés et intégrés. L'outil Rollyo, vu plus haut (voir page 37 La personnalisation), permet de se constituer son propre moteur spécialisé voire de le mettre à disposition des utilisateurs de son site web. On peut consulter un répertoire de moteurs spécialisés disponible à l'adresse suivante : <http://www.mylinea.com/moteurs-specialises>.

11. L'extension plurimédia

Les moteurs de recherche ne se limitent plus depuis longtemps à l'indexation de documents html, et proposent d'accéder à d'autres ressources : articles de presse récents, dépêches d'actualités, forums, images, mais aussi désormais vidéo et audio, avec plus ou moins de bonheur, car la pertinence de l'indexation reste discutable. On pourra ainsi tester :

- Yahoo! : <http://video.search.yahoo.com/> ; <http://audio.search.yahoo.com/>
- Google : <http://video.google.com/>
- Blinkx, qui permet de chercher dans les fichiers vidéo issus de la télévision : <http://www.blinkx.tv/>
- Truveo, moteur vidéo américain racheté par AOL qui fait beaucoup parler de lui : <http://www.truveo.com/>
- Singingfish, moteur audio et vidéo : <http://search.singingfish.com/>

Les podcasts, ces fichiers sonores ou vidéos diffusés via la technologie RSS pour une écoute ultérieure, sont également de plus en plus nombreux sur le web, notamment sur les blogs, et les moteurs suivent cette tendance (<http://podcasts.yahoo.com/> ; <http://www.podscope.com/> par exemple).

Par ailleurs, les trois grands moteurs ont annoncé leur intention de donner accès au texte intégral d'ouvrages numérisés. Google continue de multiplier, suite à des rachats de société ou partenariats, une offre de plus en plus diversifiée.

Parmi les nombreux nouveaux services proposés, on peut citer le logiciel Google Earth qui donne accès à une représentation tridimensionnelle de photos satellites.

Extrait 2 : Comment trouver des bases de données spécialisées ? (Boîte à outils)

De nombreuses bases de données sont accessibles via le web. Ces bases, dont une grande partie est en accès payant, émanent de sociétés savantes, de fournisseurs d'information économique et financière, de diffuseurs d'articles de presse, d'offices de brevets. Elles concernent soit le monde académique, en répertoriant des données scientifiques, des articles de revues, des thèses, des actes de congrès, soit le monde de l'entreprise, en donnant accès, entre autres, à des études de marché, des articles de presse économique ou des bilans.

Les bases de données sont distribuées soit par le producteur lui-même, soit regroupées au sein de serveurs qui peuvent donner accès à plusieurs centaines de bases.

Certains de ces serveurs existent depuis plus de vingt ans (c'est le cas par exemple du serveur Dialog) et sont désormais accessibles via le web, sur la base d'un abonnement payant. On peut ainsi interroger simultanément plusieurs bases de données spécialisées.

Les bases de données offrent un triple avantage.

- Des contenus à valeur ajoutée, fiables, à jour et sélectionnés en amont par des producteurs qualifiés : études de marché, rapports d'analystes financiers, développement de projets pharmaceutiques, etc.
- Un corpus structuré en champs documentaires (auteur, mots clés, date, provenance géographique, etc.) interrogeables selon une syntaxe autorisant souvent de nombreuses options de recherche, comme les équations booléennes ou la troncature, et également des possibilités de tri des résultats selon des critères sophistiqués. De plus, les bases bibliographiques proposent une indexation mots clés souvent basée sur des thésaurus ou vocabulaires contrôlés. Cela permet non seulement de mieux cibler sa recherche, mais encore de l'enrichir par l'utilisation de termes connexes ou associés.
- Des options de visualisation et d'analyse des résultats : la base de données d'informations financières Diane permettra ainsi de comparer les bilans d'entreprises d'un même secteur. Certains serveurs de bases de données proposent en outre des possibilités de veille et de diffusion ciblée d'informations.

Bases bibliographiques

Ces bases donneront accès à des données bibliographiques, c'est-à-dire la description d'articles de périodiques, études, thèses, etc. Le résumé et l'indexation des articles sont assurés par des documentalistes ou experts du domaine, et la recherche pourra se faire à partir d'un thésaurus indiquant termes associés ou rejetés. On pourra commander le document original, notamment les articles de presse, auprès d'organismes spécialisés, le plus important en France étant l'Inist (Institut de l'information scientifique et technique), basé à Nancy.

On trouvera une liste des principales bases bibliographiques dans la fiche "Comment trouver des articles de presse ?".

Bases en texte intégral

On y trouvera le texte intégral d'articles de presse, d'études de marché, ou de brevets. La recherche s'effectue sur le texte intégral, au moyen parfois de moteurs de recherche sophistiqués : Diva-Press, par exemple, utilise le moteur Intuition de Sinequa pour générer des catégorisations automatiques. Certaines bases donnent accès parfois au résumé d'articles en complément du texte intégral.

Ces bases peuvent être distribuées par des groupes de presse : c'est le cas de Science Direct qui permet l'accès au 1 800 titres du groupe Elsevier.

On trouvera une liste de bases de données en texte intégral dans la fiche "Comment trouver des articles de presse ?".

Banques d'images

Les banques d'images des grandes agences photographiques sont payantes sur le web. Néanmoins, on peut y trouver des bases d'images gratuites : attention, ces images ne sont pas toujours libres de droits.

Art resource archive	3 millions reproductions œuvres d'art, 10 000 artistes	www.artres.com
Corbis (Microsoft)	700 000 images	www.corbis.com
Getty Images		www.gettyimages.com
Artchive		www.artchive.com
Photothèque Documentation	16 500 images	http://photographie.ladocumentationfrancaise.fr/photographie/phototheque
Agence Réunion Musées nationaux	100 000 images	www.photo.rmn.fr/
Fotosearch	e 700 000 photos et vidéos libres de droit	www.fotosearch.fr

Agrégateurs et serveurs de bases de données

Les bases de données peuvent être regroupées et distribuées au sein d'un serveur unique. Parmi les serveurs commerciaux les plus célèbres, on peut citer Dialog, DataStar (ces deux serveurs faisant partie du géant de l'information électronique Thomson), ou Questel. Ces serveurs permettent une interrogation multibases et des services d'alerte sur profils.

Répertoires de bases de données

Dadi (Database directory) n'est malheureusement plus tenu à jour. On y trouve la description d'environ 700 bases de données spécialisées. (<http://urfist.univ-lyon1.fr/bd-serv.html> ou <http://dadi.enssib.fr/>)

Le guide Sinbad dresse un panorama des sociétés qui produisent, éditent et diffusent des bases de données. (<http://sinbad.univ-lyon1.fr/user/accueil.php>)

Completeplanet recense plus de 70 000 bases de données, en anglais, sur tous sujets, classées par thèmes. On ne peut malheureusement pas interroger simultanément plusieurs bases. (<http://www.completeplanet.com/>)

Il existe également des répertoires spécialisés de bases de données :

Secteur	Producteur	Adresse url
Santé	CHU Rouen	www.chu-rouen.fr/documed/tex.html
Patrimoine culturel	Ministère de la Culture	www.culture.gouv.fr/culture/bdd/bdd3.htm
Sciences humaines et sociales	Institut des Sciences de l'Homme (Lyon)	http://calame.ish-lyon.cnrs.fr/
Économie	Réseau Ecoline (Urfist Paris)	http://web.ccr.jussieu.fr/urfist/Ecoline/Ecoline.html
Droit	Réseau Jurisguide	http://jurisguide.univ-paris1.fr

L'outil Gosh me, apparu fin 2005, permet de répercuter une requête mots clés sur plus de mille moteurs, répertoires et bases de données spécialisés, classés dans 17 rubriques différentes. Il suffit de choisir de une à cinq rubriques (en cochant les cases correspondantes) et de lancer sa requête pour obtenir des liens directs vers la première page de résultats des outils sélectionnés.

Extrait 3 : Comment trouver des fils RSS (Boîte à outils)

Rappelons que le système de la syndication (voir page 18) est de plus en plus utilisé pour les sources à fort renouvellement de contenu, et bien entendu pour les blogs. De nombreux sites importants proposent aujourd'hui un grand choix de fils RSS et les médias "traditionnels" s'y mettent rapidement (voir notamment page 103 pour la presse francophone avec Alertinfo). Les fils RSS constituent désormais des outils incontournables pour la pratique de la veille, notamment sur les appels d'offres et les marchés publics (voir le site <http://www.e-marchespublics.com/rss/>). Plusieurs outils et méthodes de repérage des fils RSS se sont développés entre 2005 et début 2006.

Utiliser les moteurs généralistes

- Le moteur de recherche MSN propose des syntaxes spécifiques pour effectuer des requêtes liées à des fils RSS ou Atom.
- Yahoo! dispose de deux fonctions intéressantes. D'une part, le moteur indique, sur les résultats à des requêtes, les pages correspondantes à un fil RSS et donne un lien direct pour l'afficher en XML ou l'ajouter à sa page de Mon Yahoo! (voir page 37). D'autre part, sur la recherche avancée de Yahoo.com (non disponible sur Yahoo.fr début 2006), on peut choisir le format .xml pour le critère File Format, qui permet en fait de limiter sa recherche aux pages web qui contiennent un fil RSS dans différents formats (.xml, mais aussi .rss, .rdf, etc.).
- Exalead, comme Yahoo!, permet de sélectionner des fils RSS a posteriori d'une requête. Il suffit alors de cliquer sur le bouton "rss" présent.
- Google, s'il ne propose pas directement de fonction particulière à rss, permet des requêtes ad hoc grâce à la fonction filetype: que l'on peut décliner et éventuellement à la fonction inurl: pour le format backend.php3 non reconnu (utile pour les sites bâtis avec le système de gestion de contenu SPIP). Toutefois, l'efficacité est moyenne car les pages au format xml se multiplient aujourd'hui sur le web, et la recherche "grippe aviaire" filetype:xml générera beaucoup de "bruit". En rajoutant le mot clé rss à cette requête, et au prix d'un peu de navigation sur les pages de résultats, on peut

toutefois parvenir à ses fins. La requête "grippe aviaire" filetype:rdf OR filetype:rss OR inurl:backend.php3 donne ainsi 113 résultats. Par ailleurs, Google a récemment développé Google Reader (service gratuit mais qui requiert un compte Google), doté d'un moteur spécifique encore très décevant en ce début 2006.

Utiliser les outils spécifiques RSS

Annuaire

De nombreux annuaires dédiés aux fils RSS se sont développés sur le web, souvent d'ailleurs en lien avec un agrégateur en ligne. Notons que les systèmes de catégories apparaissent encore pauvres en regard des annuaires généralistes.

- Au niveau international, les plus importants sont Syndic8 (<http://www.syndic8.com/>), un leader avec plus de 450 000 fils dont 16 600 créés ("scrapped" en anglais) après proposition des membres enregistrés et validation. Syndic 8 n'est toutefois pas le plus simple d'usage ; NewsIsFree (<http://www.newsisfree.com/sources/>) qui offre une recherche par catégorie, et aussi par mots clés sur le titre et la description ; Rss Network (<http://www.rssnetwork.com/>) qui propose un système développé de catégories ; 2rss.com (<http://www.2rss.com/>) ; Feed Directory (<http://www.feed-directory.com/>) ; Rssfeeds (<http://www.rssfeeds.com/>).

- Sur les sources francophones, citons : Pretty-RSS (http://pretty-rss.snyke.com/Annuaire_RSS/Annuaire_RSS.html) ; Retronimo (<http://www.retronimo.com/>) ; RssReporter (<http://www.rssreporter.net/>) ; Lamooche (http://www.lamooche.com/annuaire_rss.php) ; et aussi le récent Mon-actu.com (<http://mon-actu.com/>).

Moteurs

Pour les moteurs, Feedster (<http://www.feedster.com/>) apparaît comme le leader avec plus de 21 millions de flux indexés (soit 150 millions de notes depuis mars 2003), une recherche qui peut se limiter aux blogs ou à un fil spécifique.

Voir aussi EasyRSS, qui a l'ambition de devenir le premier moteur européen de fils RSS ; Rss Network, vu plus haut mais qui joue aussi le rôle de moteur (cocher "Search items" sur la page de résultats à une requête).

On voit aussi apparaître des outils thématiques, souvent des agrégateurs, comme Medlogs.com pour le domaine médical ou des flux multi-sources proposés par des internautes experts : voir par exemple Sedna RSS de Service Doc.info (Stéphane Cottin, <http://www.servicedoc.info/sedna/>) pour la documentation et le droit, ou bien Pipologie pour l'actualité des moteurs de recherche, ou Strategic Board pour l'actualité des technologies de l'information (en anglais, plus spécialisé sur les blogs, <http://www.strategicboard.com/>).

Ces sources s'avèrent fort utiles pour identifier les flux importants d'un secteur.

Utiliser les agrégateurs privilégiant la navigation sociale

Un agrégateur en ligne tel RSS4You (<http://www.rss4you.com/>) met le focus sur le partage collaboratif de fils RSS : avec un système de vote et le partage des flux RSS inscrits, il permet à ses membres de mieux cibler les autres membres qui ont les mêmes intérêts.

Bloglines, le lecteur RSS en ligne appartenant désormais à Ask Jeeves, permet le même type de partage (sans les votes) et donc, de recherche de nouveaux flux sur le principe de la navigation sociale. L'outil permet aussi d'importer entièrement la sélection publique d'un utilisateur au format OPML.

Fiche technique de l'ouvrage "Net Recherche" :

Auteurs : Armelle Thomas - Véronique Mesguich
Editeur : ADBS
2006 - 148 pages - ISBN : 2-84365-083-6
Prix : 23 € TTC

Sommaire :

Diversité des besoins, diversité des outils - Moteurs de recherche : pour en finir avec les idées reçues - Les métamoteurs : entre traditions et innovations - Pour une recherche thématique : des annuaires généralistes aux portails spécialisés - L'émergence du web "social" - En guise de conclusion... - Boîte à outils : Fiches pratiques Questions-Réponses / Panorama des outils de recherche généralistes et spécialisés.

Le sommaire complet (format PDF) :

http://www.adbs.fr/uploads/ouvrages/4407_fr.pdf

Commander le livre : <http://www.adbs.fr/site/publications/ouvrages/104.php>