

Net recherche : Comment trouver des articles de presse ?

[Retour au sommaire de la lettre](#)

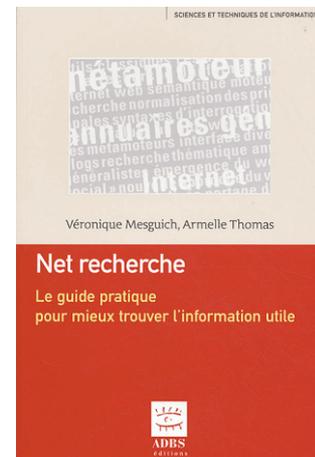
Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Nous vous proposons dans cet article une "fiche pratique" extraite du livre "Net recherche" qui vient de sortir aux éditions ADBS. Elle nous décrit les grands domaines des sources d'informations journalistiques présentes sur le Web : revues universitaires, bases bibliographiques, agrégateurs, archives ou outils de journalisme citoyen, et nous propose quelques pistes de réflexion et d'action pour trouver les meilleurs articles sur un sujet donné en un minimum de temps....

Le livre "**Net recherche 2009 : le guide pratique pour mieux trouver l'information utile et surveiller le web**" vient de paraître aux éditions ADBS (<http://www.adbs.fr/net-recherche-2009-le-guide-pratique-pour-mieux-trouver-l-information-utile-et-surveiller-le-web-61812.htm>). En voici une courte présentation proposée par l'éditeur :

Sous l'apparente facilité d'utilisation des moteurs de recherche se cache en effet une réalité complexe, et le secret de la réussite d'une recherche ou d'une veille passe autant par la maîtrise des aspects techniques que par la capacité à évaluer et sélectionner les sources pertinentes.

En cette troisième édition profondément renouvelée, développée et mise à jour, Net recherche vise à offrir à toute personne amenée à effectuer des recherches sur Internet un panorama des outils et méthodes existant à ce jour, en intégrant les dispositifs qui permettent de surveiller le web à moindre coût. Cet ouvrage s'adresse notamment aux professionnels de l'information (documentalistes, bibliothécaires, veilleurs), aux enseignants, chercheurs et étudiants, et à tous les autres "travailleurs du savoir" confrontés à la complexité croissante et à l'inflation de l'information en ligne. Ils y découvriront des conseils méthodologiques mais aussi nombre de "trucs et astuces" destinés à optimiser le processus de recherche ou de veille, des informations précises sur les évolutions en cours, ainsi que des présentations d'outils et des listes d'adresses utiles.



Les auteurs ont demandé à Olivier Andrieu, éditeur du site Abondance.com d'en signer la préface. En accord avec les auteurs et l'éditeur, nous vous présenterons, pendant quelques mois, quelques "fiches pratiques" présentes dans ce livre. Nous espérons qu'elles vous permettront de mieux chercher l'information sur le Web... Après "*Comment choisir ses mots-clés ?*" et "*Comment effectuer un sourcing de qualité*" le mois dernier, nous nous intéressons ce mois-ci à la presse en ligne...

Fiche 3. Comment trouver des articles de presse ?

C'est désormais une évidence : le développement d'Internet a bouleversé le monde des médias, et plus particulièrement la presse écrite. Selon une étude réalisée en décembre 2008 par le Pew Center, Internet est devenu la deuxième source d'information des Américains, derrière la télévision mais devant, et pour la première fois, la presse écrite (<http://pewresearch.org/pubs/1066/internet-overtakes-newspapers-as-news-source>). À l'instar de nombreux métiers de médiation, la profession de journaliste fait l'objet d'une mutation profonde, tant en matière de méthodes d'investigation – l'accès aux sources d'information s'étant beaucoup démocratisé – que de diffusion. Selon Pierre Chappaz, fondateur du site Wikio, la distinction entre blogueur et journaliste est devenue floue : "Le mur de Berlin qui sépare journalistes et blogueurs est déjà tombé", écrivait-il sur son blog en juin 2008 (<http://www.kelblog.com/2008/06/les-journaliste.html>).

Pour les médias traditionnels, le support web permet par ailleurs d'introduire de nouveaux services et plus d'interactivité : forums, newsletters, blogs, commentaires, sondages, diffusion d'information continue, dossiers d'actualité, voire filtrage et diffusion d'information ciblée, etc. Autre effet de l'"immédiateté" du support Internet : des magazines hebdomadaires peuvent diffuser sur leur site web des dépêches réactualisées plusieurs fois dans la journée.

Enfin, de nouvelles formes de regroupement de plusieurs sources d'actualité sont apparues récemment, à travers les agrégateurs de presse, et plus récemment les fils RSS.

Certains rares journaux restent absents de la Toile et n'existent que sur support papier. A *contrario*, de nombreux magazines ou newsletters se sont créés sur le web et ne sont disponibles que sous forme électronique : l'année 2007 a vu la création de plusieurs sites de "journalisme citoyen" (voir plus loin). Le mouvement mondial des archives ouvertes, c'est-à-dire le dépôt, par les chercheurs, de leurs publications récentes et leur stockage sur des serveurs en libre accès, a pris une grande ampleur depuis plusieurs années.

La situation des médias sur le web est ainsi particulièrement complexe et regroupe des cas de figure très différents : depuis les revues universitaires à comité de lecture jusqu'aux nouveaux "médias sociaux" tel Twitter, on trouve sur le web une palette infinie de contenus en toutes langues, de formules d'accès, de tarification, etc.

Les sites de presse quotidienne générale et régionale

Les quotidiens nationaux ou régionaux disposent depuis longtemps de sites, dont certains sont devenus de véritables portails spécialisés. Le Monde offre par exemple l'accès à des dossiers spécialisés et permet d'héberger des blogs ; le site des Échos diffuse des chaînes thématiques consacrées à des secteurs d'activité ; de nombreux quotidiens régionaux permettent de consulter des informations locales. Si la recherche par mots-clés dans le texte intégral des archives est généralement gratuite, la consultation des articles, en revanche, fait souvent l'objet d'un paiement qui s'effectue soit à l'acte, par carte bancaire, soit sur abonnement.

Certains éditeurs ont pris le parti de rendre leurs archives gratuites sur la Toile, dans le but d'attirer les internautes *via* les moteurs de recherche, et de développer leurs recettes publicitaires. C'est le cas entre autres, aux États-Unis, du New York Times et en France du quotidien financier La Tribune depuis avril 2008.

Plusieurs annuaires de presse généraliste permettent d'identifier aisément des sites de quotidiens nationaux ou régionaux dans le monde entier.

- **Gigapresse** : plus de 800 journaux francophones classés par thème, et 10 000 journaux internationaux, classés par pays ;
- **Newspapers** : presse mondiale, généraliste et spécialisée ;
- **Bibliothèque virtuelle des périodiques** : projet franco-québécois, répertoire de 650 revues et magazines électroniques offrant leur contenu sur Internet.

L'excellent répertoire de presse mondiale **Webdopresse**, créé sur le site du magazine suisse L'Hebdo, n'est malheureusement plus maintenu depuis décembre 2008.

La presse professionnelle

La presse professionnelle émane souvent de petits éditeurs plutôt que de grands groupes de médias, et est de ce fait moins représentée sur le web que la presse généraliste. On trouve en ligne, et parfois gratuitement, le contenu complet de revues et magazines dans certains secteurs, notamment liés aux nouvelles technologies (informatique, multimédia) ; c'est moins le cas pour d'autres secteurs plus traditionnels. De nombreuses revues et magazines spécialisés disposent néanmoins d'un site web, l'accès aux archives restant encore parfois payant, ou réservé aux abonnés à la version papier. Afin d'identifier des sites français de presse professionnelle ou spécialisée, on peut consulter des répertoires comme Tarif media ou l'annuaire des adhérents de la Fédération nationale de la presse spécialisée (FNPS). Pour la presse spécialisée mondiale, utiliser l'**Internet Directory of Publications**, répertoire de plus de 150 000 publications.

Les revues universitaires

L'adage "Publish or perish", selon lequel les chercheurs se doivent de publier leurs résultats pour acquérir une visibilité et progresser dans leur carrière, s'illustre tout particulièrement sur le web. Pour autant, les pratiques peuvent différer d'un éditeur à l'autre, et d'une discipline à l'autre. Certaines revues, notamment dans le domaine des sciences sociales, peuvent être disponibles gratuitement en texte intégral sur le web. L'édition de revues scientifiques et techniques est dominée par quelques grands groupes d'édition comme Springer ou Elsevier. Ces éditeurs proposent généralement un abonnement payant à la version électronique de leurs revues. On peut avoir accès aux sommaires et parfois aux résumés des articles via le site web des éditeurs.

Pour retrouver des articles publiés par des chercheurs en sciences pures, économiques ou sociales, on peut utiliser plusieurs modes d'accès.

Bases bibliographiques

On y trouve des références, accompagnées de résumés, d'articles de périodiques. Le contenu est en grande partie en langue anglaise.

Principales bases bibliographiques d'articles de périodiques :

Ingenta www.ingenta.com	Gratuit (commande d'articles payante)	Multidisciplinaire	25 millions d'articles issus de plus de 30 000 publications
Inist artscienciences.inist.fr	Gratuit (commande d'articles payante)	Scientifique (dont les articles de la base Pascal, mais sans recherche fine par mots-clés)	Plus de 11 millions d'articles
Science Direct www.sciencedirect.com	Payant	Scientifique et médical dont 7 millions d'articles en texte intégral	Plusieurs millions d'articles
Medline (Pubmed) www.ncbi.nlm.nih.gov	Gratuit	Médical	17 millions d'articles
Pascal ingenierie.inist.fr/article27.html	Payant	Scientifique	17,4 millions d'articles issus de plus de 3 000 revues
Inspec (Institute of Electronic Engineers) www.theiet.org/publishing/	Payant	Electronique	10 millions d'articles
Chemical Abstracts (American Chemical Society) www.cas.org	Payant	Chimie	30 millions d'articles issus de 10 000 revues

Sites des chercheurs

Certains chercheurs publient leurs propres articles, en texte intégral et gratuitement, sur leur site ou leur blog. On peut ainsi parfois y retrouver le texte intégral de certains articles référencés dans les bases bibliographiques.

Portails de revues universitaires

Il existe plusieurs portails répertoriant des revues et donnant, le cas échéant, des accès gratuits aux contenus. Les sciences humaines sont particulièrement bien représentées à travers plusieurs "bouquets" de revues francophones.

Cairn est un projet mixte, public et privé, qui rassemble plusieurs éditeurs et partenaires institutionnels, dont la BnF. Certains articles sont en consultation gratuite, mais la grande majorité est accessible sur abonnement ou en "pay per view". Début 2009, Cairn regroupait plus de 50 000 articles issus de 196 revues en économie, droit, sciences humaines et sociales. On peut créer un compte afin d'automatiser des alertes sur des parutions ou sauvegarder sa bibliographie.

Persée est un portail de revues scientifiques en sciences humaines et sociales, créé par le ministère de l'Éducation nationale, qui a pour vocation la numérisation et la mise en ligne des collections rétrospectives de ce corpus.

Erudit.org rassemble une collection d'une quarantaine de revues spécialisées dans diverses disciplines des sciences sociales.

Revues.org est le plus ancien portail français de revues de sciences humaines et sociales. Il a pour partenaires le Cnrs, l'EHESS et le TGE Adonis. Cairn et revues.org ont conclu récemment un partenariat.

JSTOR (Journal Storage) enfin, fondé en 1995, donne accès au texte intégral de plus de 700 revues académiques américaines ou anglaises en sciences sociales et économiques, représentant plus d'1,5 millions d'articles mais qui correspondent plutôt à des collections rétrospectives qu'à des abonnements courants.

Le moteur **In-Extensio** indexe plus de 200 000 documents en sciences sociales, issus de plusieurs de ces portails (Persée, Cairn, Erudit, Revues.org) mais également de Gallica et du serveur d'archives ouvertes HAL-SHS (voir plus loin).

Moteurs de recherche spécialisés

Certains outils de recherche se limitent à un corpus spécialisé de sites universitaires et de bases bibliographiques. Parmi les plus importants, on peut citer les deux suivants.

Scirus : cet outil de recherche spécialisé, basé sur le moteur Fast, indexe en profondeur des sites à caractère scientifique sélectionnés en amont, ainsi que des bases de données gratuites spécialisées en sciences. Son point fort : la possibilité d'affiner sa recherche à partir de mots-clés suggérés automatiquement par le moteur, ou de rechercher sur des dates précises. Scirus a été mis au point par l'éditeur Elsevier.

Google Scholar : lancé début 2005 par Google, ce module est destiné essentiellement à la communauté des chercheurs et des étudiants. Il collecte des contenus à caractère académique : publications universitaires, articles scientifiques, thèses, ouvrages, rapports de recherche, etc. La pertinence est basée sur une recherche dans le texte intégral (même si celui-ci n'est pas en accès gratuit sur le web) et sur le nombre de fois où l'article est cité. L'analyse des citations fait l'objet d'un lien à part ("*cited by*"). Par contre, l'information concernant les éditeurs est moins clairement indiquée que dans Scirus, et la recherche avancée ne permet ni d'affiner par mots-clés ni de rechercher sur des dates précises. Google Scholar "moissonne" également le catalogue commun des bibliothèques universitaires françaises (Sudoc), des catalogues de bibliothèques ainsi que des dépôts d'archives ouvertes.

Les revues électroniques ou newsletters

De nouvelles formes de publication sont apparues, sous forme uniquement électronique, sans équivalent papier. On les appelle *newsletters* ou *webzines*. On peut y trouver, souvent gratuitement, de l'information exclusive et précieuse, à caractère anticipatif, qui pourra se révéler utile pour la pratique de la veille. Attention cependant : certaines newsletters ne sont que de simples supports commerciaux ou publicitaires. À l'heure du web 2.0, l'explosion des blogs et flux RSS n'a sans doute pas favorisé l'émergence récente de nouvelles lettres d'information. Toutefois, les éditeurs continuent souvent à "doubler" leurs flux RSS d'une

newsletter, car certains internautes soit ne sont pas encore familiers avec cette technologie, soit préfèrent le mail classique.

Les newsletters, échappant par définition au circuit classique de l'édition, sont difficiles à identifier sur la Toile. L'annuaire des lettres d'information, créé par le site **Pidinfo**, n'existe plus, de même que d'autres répertoires de newsletters mentionnés dans les précédentes éditions de cet ouvrage. Le moyen le plus efficace pour repérer une newsletter spécialisée reste le suivant : par exemple, dans le domaine du comportement animal, il suffira de saisir, via un moteur de recherche classique, les mots-clés *newsletter* et "*animal behavior*".

Les agrégateurs de presse et bases de données d'articles

Ces outils permettent de rechercher simultanément par mots-clés dans les archives en texte intégral de plusieurs quotidiens nationaux ou régionaux ou de revues spécialisées, et ce sur plusieurs années. Ces services sont accessibles généralement sur abonnement. Certains d'entre eux (Europresse, Pressed) permettent d'interroger gratuitement la base, la consultation des articles restant payante. Parmi les principaux agrégateurs de presse, on peut citer :

- **Lexis Nexis** : presse internationale ;
- **Factiva** : presse internationale ;
- **Europresse** : presse francophone, accès avec ou sans abonnement ;
- **Pressed** : presse francophone, bonne couverture régionale.

Il convient de mentionner le service "Actualités" de Google qui, en février 2009, rassemble 25 000 sources d'information mises à jour en continu. Il ne s'agit pas d'une base de données : les articles ne sont pas archivés mais font simplement l'objet d'un lien vers leur site d'origine. Ce service est donc davantage adapté pour une recherche sur l'actualité immédiate ou pour une veille par mots-clés (option "*E-mail alerte*" ou désormais diffusion par flux RSS ou Atom) que pour une recherche rétrospective sur des articles plus anciens.

Yahoo! propose aussi un service d'actualités, le nombre de sources accessibles étant inférieur à celui de Google Actualités. Le service Newsbot, créé par MSN, ne semble plus être disponible. N'oublions pas non plus la base de données **Findarticles** qui permet de consulter gratuitement le contenu de 10 millions d'articles, issus pour la plupart de revues américaines. Ce site, revendu en 2007 par Looksmart, a pris une orientation très grand public, mais conserve néanmoins une rubrique "*Business Library*".

Si d'anciens services comme **NewsTrawler** ou **NewsIndex** ne sont plus disponibles, des agrégateurs "historiques" perdurent en revanche, comme le britannique **Newsnow**, présent sur le web depuis 1997. Les articles, en provenance de 34 890 sources annoncées, sont triés sur la base de quelque 600 catégories regroupées dans des grandes thématiques. La version gratuite ne permet qu'une recherche avec un seul mot-clé mais les différentes versions payantes ajoutent diverses fonctionnalités (tels des flux RSS).

Deux nouveaux venus francophones en 2006 dans le monde des agrégateurs méritent vraiment le détour.

- **Wikio** surfe sur la vague web 2.0 en surveillant aujourd'hui près de 60 000 médias et blogs francophones en temps réel (contre 34 000 en 2007 !) et classe les documents par thématiques (tags). Les internautes ont la possibilité d'infléchir le classement des articles en votant pour ceux qui leur paraissent les plus intéressants ; on peut également apposer des commentaires à la suite de chaque article. Un service de surveillance mots-clés est disponible sous forme de flux personnalisé. Wikio propose également un classement des blogs francophones les plus influents.

- **Synthesio**, lancé par une jeune start-up française, est plus orienté business, et propose d'ailleurs un abonnement payant pour surveiller des médias classiques, blogs et forums. L'interface de ce nouvel agrégateur apparaît à l'usage particulièrement ergonomique et intuitive, avec la possibilité de filtrer directement les résultats par date, pays, type d'information (blog, actualité, droit, étude, événement, etc.), type de sources (site spécialisé, généraliste, entreprise, presse nationale, institution, etc.), langues et sources. Toutes les sources sont préalablement décrites manuellement par des professionnels de l'information

(notamment 700 secteurs) et les données peuvent alors s'organiser automatiquement en fonction de leur provenance.

Notons la facilité de créer un moteur personnalisé à partir d'un choix de sources, et d'automatiser la création de newsletters à partir des articles retenus. La technologie Synthesio est utilisée entre autres par l'agrégateur Widepress.

Par ailleurs, l'agrégateur international **Newstin** permet de surveiller les dépêches du monde entier et d'effectuer des traductions automatiques. On peut personnaliser des requêtes et créer des flux personnalisés. Voir l'innovant **Silobreaker** (filtres, extraction automatique de concepts, etc.).

Les fils RSS

On peut désormais composer soi-même des "bouquets" de sources d'actualité en syndiquant plusieurs fils d'information à l'intérieur d'un lecteur RSS. C'est le principe de la syndication. Le lecteur RSS **Alertinfo** permet ainsi de fédérer les fils d'information issus de plusieurs grands quotidiens francophones (Le Monde, Les Échos, Le Figaro, La Tribune, etc.). Ce lecteur peut être téléchargé gratuitement depuis le site du groupement professionnel Geste. Il faut filtrer judicieusement les contenus (en créant, par exemple, des filtres par mots-clés ou domaines) sous peine d'être "inondé" par le flux d'information ainsi généré.

D'autres agrégateurs proposent de suivre l'actualité dans des domaines ciblés grâce à une présélection en amont de flux RSS. Par exemple, le très utile **Eufeed** centralise les flux de plus de mille journaux de tous les pays européens !

Les archives ouvertes et sites de prépublications

De nombreuses initiatives ont été menées pour permettre la mise à disposition des résultats de la recherche au plus grand nombre, *via* les réseaux. Le déclaration de Berlin, en 2003, a constitué une étape majeure dans ce mouvement du libre accès (*Open Access Initiative*). À travers ce texte, les principaux organismes de recherche européens s'engagent à "*promouvoir un Internet qui soit un instrument fonctionnel au service d'une base de connaissance globale et de la pensée humaine*".

Les chercheurs avaient depuis longtemps la possibilité de publier des articles avant parution dans des revues à comité de lecture. Il existe en effet de nombreux sites de prépublication d'articles, la plupart en langue anglaise. La base **Arxiv**, base de preprints du Laboratoire de Los Alamos, précurseur des archives ouvertes, fut l'une des pionnières du domaine. Spécialisée en mathématiques et physique, cette base continue de s'enrichir chaque mois de 4 000 nouveaux documents. Les sites de prépublications offrent une grande souplesse en permettant à l'auteur d'auto-archiver, et même de mettre à jour en ligne, ses documents. Les versions successives sont disponibles sur le serveur.

Ce modèle a été repris par le Centre pour la communication scientifique directe (**CCSD**), unité du CNRS créée en 2000. Le CCSD maintient le serveur Hal destiné à la publication de documents scientifiques de toutes disciplines. On trouvera également une liste d'institutions américaines diffusant des preprints ou des rapports techniques en libre accès sur le site **Virtual Technical Reports Center**.

De nombreuses archives ouvertes respectent le protocole OAI (*Open Archives Initiative*) facilitant l'échange de données. Ce protocole permet de créer, d'alimenter et de tenir à jour, par des procédures automatisées, des réservoirs d'enregistrements qui signalent, décrivent et rendent accessibles des documents, sans les dupliquer ni modifier leur localisation d'origine.

Plos (Public Library of Science) est une organisation à but non lucratif dont la mission est de donner libre accès à l'information scientifique et technique, médicale et biologique mondiale. Le portail Plos Journals met à disposition, en *open access*, des journaux scientifiques et techniques : l'accès au texte intégral est gratuit, les coûts de publication sont pris en charge par l'auteur ou l'organisme dont il dépend, la version papier de la revue restant payante au prix de l'impression.

Le site **Oaister**, lancé en 2005 par la Bibliothèque de l'Université du Michigan, offre à travers un outil unique l'interrogation de plusieurs centaines de sites d'archives institutionnelles, respectant la norme OAI-PMH. Il compte début 2009 près de vingt millions de documents (le nombre a doublé en deux ans !) provenant de 1 072 éditeurs. L'objectif est de créer une collection librement accessible de documents académiques.

L'accès libre peut concerner également des documents pédagogiques : c'est le cas par exemple de l'**Université numérique Ingénierie et Technologie**, qui donne accès à des cours en ligne dans le domaine des sciences de l'ingénieur. Ce site utilise le logiciel ORI-OAI (Outil de référencement et d'indexation, réseau de portails OAI).

On n'oubliera pas le moteur spécialisé en sciences sociales **In-Extenso**, cité plus haut, ainsi que le répertoire **DOAJ (Directory of Open Access Journals)** des revues gratuites et disponibles en libre accès sur la Toile. Début 2009, ce répertoire comptait plus de 3 800 revues en libre accès, classées par grand thème. Cet annuaire, publié par l'université de Lund en Suède, est devenu un outil incontournable pour le repérage des revues scientifiques publiées en libre accès.

Afin de suivre l'actualité très riche du mouvement du libre accès, on peut consulter le blog **Libre accès à l'IST** créé sur le site de l'Inist.

Le "journalisme citoyen"

Terminons enfin ce tour d'horizon des médias du web en évoquant la notion de "journalisme citoyen". L'un des pionniers dans le domaine a été **Agoravox**, lancé dès 2005 par Carlo Revelli et Joël de Rosnay.

Le principe est de permettre à des journalistes "non professionnels", ou à tout citoyen, de s'exprimer ou commenter des faits d'actualité avec un regard différent de celui des médias traditionnels. Les articles doivent toutefois correspondre à une politique éditoriale. D'autres projets journalistiques indépendants se sont développés plus récemment, à l'initiative d'anciens journalistes de Libération (**Rue 89** avec Pierre Haski et Pascal Riché) ou du Monde (**Mediapart**, lancé en 2007 entre autres par Edwy Plenel). Ils se positionnent parfois sur "la réflexion critique sur les médias", comme le site **Arrêt sur image**, créé par le journaliste Daniel Schneidermann en septembre 2007 à la suite de l'arrêt de l'émission du même nom sur France 5.

Un univers Netvibes, intitulé **Portail des actualités** intègre d'ailleurs un onglet "Journalisme citoyen" (http://www.netvibes.com/actus-journalisme_citoyen).

Véronique Mesguich et Armelle Thomas

Auteurs du livre "Net recherche 2009 : le guide pratique pour mieux trouver l'information utile et surveiller le web". Extraits du chapitre 7.

Réagissez à cet article sur le blog des abonnés d'Abondance :

<http://abonnes.abondance.com/blogpro/2009/06/net-recherche-comment-trouver-des.html>