

Comment exploiter les moteurs de recherche d'actualités ?

[Retour au sommaire de la lettre](#)

Apparus il y a quelques années, les moteurs de recherche d'actualités se sont rapidement imposés comme l'un des piliers du monde de la recherche, au même titre que la recherche de sites Web ou la recherche d'images. La plupart des grands moteurs traitent désormais les "news" et il existe une multitude d'outils spécialisés. Alors même que ces outils verticaux drainent de plus en plus de visites, les possibilités offertes pour y référencer des contenus restent cependant relativement méconnues... Nous vous offrons ici un panorama des outils disponibles et un aperçu des facteurs à prendre en compte pour figurer en bonne place dans les listes de résultats !

La fréquentation des moteurs de recherche d'actualités a connu une forte progression dernièrement. Aux Etats-Unis, le cabinet Nielsen//NetRatings a annoncé, en janvier 2005, que le nombre de lecteurs de Yahoo! News était désormais supérieur à celui de CNN, avec environ 23 millions de visiteurs par mois. En juillet de cette même année, ce cabinet plaçait Yahoo! News en tête des sites d'actualités US avec 23,2 millions de visiteurs uniques, devant AOL News (16,5 millions), Google News (6,8 millions) et Topix.net (2,4 millions).

En décembre dernier, Nielsen//NetRatings a par ailleurs estimé que la recherche d'actualités occupe une place de plus en plus importante dans le nombre total de recherches effectuées en ligne sur les moteurs de recherche, avec une progression de 12,7% du nombre de requêtes effectuées entre juin et octobre 2005.

Preuve supplémentaire de l'intérêt qu'ils suscitent, au moins quatre moteurs de recherche d'actualités spécialisés ont été rachetés durant ces deux dernières années. Le service français Net2One a tout d'abord été revendu en 2004 au prestataire de veille Presse+ (qui a depuis lui-même été racheté par le groupe britannique TNS Secodip). Cette même année, Rocket Technologies, autre spécialiste d'origine canadienne responsable de RocketNews, a lui été repris par l'américain Zeppelin Energy (groupe œuvrant à la fois dans l'informatique et le secteur pétrolier). En octobre 2005, l'agrégateur de contenus Moreover est quant à lui tombé dans l'escarcelle de VeriSign, un spécialiste des noms de domaine, pour 30 millions de dollars. Enfin, dernièrement, trois éditeurs de journaux américains (Tribune, Knight-Ridder et Gannett) se sont unis pour prendre une participation de 75% dans Topix.net.

Chronologie des principaux lancements

Il existe des centaines d'outils, dont une grande majorité sont anglophones. Ils permettent de rechercher des actualités en ligne diffusées par toutes sortes de médias, qu'il s'agisse de journaux locaux, de quotidiens nationaux, de sites de stations de radio, de chaînes de télévision ou encore, mais moins souvent, de weblogs). Nous nous concentrons ici sur les principaux d'entre eux :

Année	Toujours actif	Moteur de recherche d'actualités	Langue(s)
1995	Oui	Yahoo! News	UK / FR / ES...
1996	Oui	News Index	UK
1997	Oui	Excite NewsTracker	UK
	Oui	Net2One	FR / UK
	Oui	TotalNEWS	UK
1998	Oui	Moreover	UK
1999	Oui	NewsTrove	UK
	Non	Northern Light Special Collection	UK
2000	Oui	LookSmart FindArticles	UK
	Oui	Ananova (Orange UK)	UK
2001	Non	Altavista News	UK

Année	Toujours actif	Moteur de recherche d'actualités	Langue(s)
	Oui	Ask Jeeves News	UK
	Oui	Daypop	UK
	Oui	RocketNews	UK
2002	Oui	DeepIndex News	FR
	Oui	Google News	UK / FR / ES...
	Oui	Topix.net	UK
2003	Non	Allzenews	FR / UK
	Oui	MSN Newsbot	UK / FR / ES...
2004	Oui	Oscop	FR
2005	Oui	Feedster	UK

Fonctionnement technique des moteurs de recherche d'actualités

Pour constituer leurs index, les moteurs de recherche d'actualités ont plusieurs possibilités. La première consiste à utiliser des spiders "classiques" parcourant et indexant périodiquement une sélection de sites d'actualités. Google utilise cette méthode, comme l'explique Nathan Stole (le directeur produit de Google News) : *"Google News a ses propres crawlers qui scannent très, très rapidement tous les nouveaux sites que nous avons dans notre corpus, essayant d'identifier les articles dès qu'ils apparaissent"* avant d'en *"extraire le titre, le corps du texte et les photos"* (<http://searchenginewatch.com/searchday/article.php/3556766>).

Une deuxième méthode, plus répandue semble-t-il, consiste à indexer des flux XML, transmis par les sites sur une base régulière. Avantage de cette méthode : le moteur considère parfois qu'il s'expose moins à des réactions "hostiles" de la part d'éditeurs ne voulant pas que leurs contenus soient référencés et utilisés car les fils RSS sont par nature voués à l'export d'information.

Comme leurs grands frères spécialisés sur les sites Web, les moteurs de recherche d'actualités utilisent en outre un moteur d'indexation pour extraire les mots-clés des textes publiés. Les différentes informations collectées sont stockées dans une base de données. A noter que la plupart des moteurs de recherche "généralistes" stockent ces actualités dans des bases séparées (d'où la nécessité de mener un référencement spécifique pour les articles).

Une fois cette étape passée, les outils qui classent les articles dans des catégories thématiques "clusterisent" les résultats. Pour cela, ils analysent chaque terme employé dans les articles et effectuent des regroupements entre les articles qui ont en commun l'utilisation de certaines expressions (<http://www.seroundtable.com/archives/002914.html#trackbacks>).

A chaque interrogation, un système de *ranking* classe les résultats en fonction d'algorithmes spécifiques au moteur. Contrairement à la situation qui prévaut sur les moteurs Web, il est à noter que le PageRank (ou le nombre et la qualité des liens pointant vers une actualité) est ici peu pris en compte comme critère de pertinence. Ce système n'est en effet guère approprié pour déterminer la pertinence d'une news, qui est souvent d'autant plus intéressante qu'elle est récente et a encore peu de liens pointant vers elle sur le Web. En revanche, les critères les plus répandus pour prioriser les résultats sont la "fraîcheur" de l'actualité et l'importance de la source la publiant.

De nombreux outils de recherche spécialisés sur les actualités associent aussi des vignettes à leurs résultats. Ils ont pour cela recours à une application permettant de traiter les images associées à un article ou – lorsque l'article ne comporte aucune photo – d'analyser toutes les images d'un site pour y sélectionner l'illustration la plus pertinente possible.

Modèles économiques

Pour générer des revenus, les moteurs de news ont plusieurs possibilités :

- La **vente d'espaces publicitaires**. Les moteurs WorldNews et Topix.net offrent, entre autres, cette possibilité.
- L'**intégration de liens sponsorisés dans la liste de résultats**. Oscope, Topix.net et Rocketinfo incluent des publicités AdSense de Google dans leurs listes de résultats.
- L'**intégration de liens sponsorisés dans les fils XML**. Depuis 2004, Moreover intègre des liens sponsorisés fournis par le prestataire Kanoodle à ses fils RSS.
- Le **positionnement payant** de résultats. Fin 2005, le New York Times a, par exemple, signé un contrat avec Topix.net, pour placer certains titres en tête de plusieurs catégories de ce moteur.
- La **vente de résultats et de fils thématiques à d'autres sites ou moteurs**. Topix.net propose ses résultats à des FAI américains et à des moteurs comme Ask Jeeves.
- La **vente de licences logicielles**. Rocketinfo commercialise une solution de recherche payante sur la cible des entreprises. Le britannique NewsNow offre également des applications payantes de veille sur l'actualité.

Relations entre les moteurs de news et les éditeurs de sites

A noter que l'éventuelle monétisation des contenus des moteurs de recherche d'actualités - et tout particulièrement de celui de Google - a entraîné plusieurs levées de boucliers de la part des éditeurs français.

En 2003 et en 2004, une dizaine d'entre eux ont, avec l'appui du Geste (voir entretien de Philippe Jannet dans cette lettre), demandé à ce que leurs contenus ne soient plus référencés dans Google News. Ils craignent vraisemblablement de voir ce dernier profiter de leurs contenus pour générer des revenus publicitaires.

Signe du malaise de certains éditeurs par rapport à la finalité du service, Jean Rebuffat (chef de la rédaction du quotidien belge le Soir) déclarait en 2003 au Journal du Net : *" Il n'y a que pour les photos que je suis réticent, sinon cela me semble acceptable puisque le droit de citation existe. Tant que Google News reste un vitrine du savoir technologique de Google et qu'il n'y a pas de modèle économique qui y est associé, cela ne me semble pas incorrect"* (<http://www.journaldunet.com/0310/031022google.shtml>).

Début 2005, l'Agence France Presse (AFP) a quant à elle attaqué Google en justice pour violation de copyright sur des images et des informations lui appartenant. Google News reprenait alors en partie des textes et des images de l'agence sans son accord (ces contenus ont depuis été retirés par Google).

AlertInfo

En mai 2005, le GESTE (Groupement des éditeurs de services en ligne) a lancé AlertInfo (<http://www.alertinfo.fr/>), un lecteur RSS gratuit agrégeant des fils de contenus en provenance de sites de périodiques en ligne français. 400 fils RSS sont intégrés par défaut parmi lesquels les fils d'information des Echos, de Libération, du Monde ou de La Tribune. Pour intégrer un fil, il faut être membre du Geste, et obtenir l'aval de son conseil d'administration.

Ce lecteur est une adaptation francophone du service RSS open source FeedReader. Il a été développé par les équipes techniques du site LesEchos.fr en collaboration avec une quinzaine d'autres éditeurs de presse en ligne (La Tribune, Le Monde, Le Figaro, L'Entreprise, L'Express, L'Expansion, France 2, Libération, RTL, Le Nouvel Observateur, France 3, L'Equipe, 01Net, ZDNet et Businessmobile).

Une fois installé, le logiciel AlertInfo offre un accès rapide à de nombreux fils RSS organisés par thématiques (14 au total).

Typologie des moteurs de recherche d'actualités

Dès le milieu des années 90, des moteurs tels que Yahoo! ont commencé à proposer une sélection de news aux internautes.

Aucun moteur ne couvre, cependant, l'étendue des sources disponibles. Pour une recherche exhaustive, il est donc fortement conseillé d'utiliser en parallèle plusieurs outils !

Moteurs de recherche incluant des fonctionnalités de recherche sur les news

Après une première version "embryonnaire" lancée en 1995, Yahoo! News (<http://fr.news.yahoo.com/>) fait désormais office de précurseur sur le marché de la recherche d'actualités dans les pays anglo-saxons. En France, ce n'est qu'en 2004 que Yahoo! a ajouté un nombre conséquent de titres de la presse française à son service. Il proposait avant cette date essentiellement des dépêches d'agences de presse (AFP, AP et Reuters). Google a, pour sa part, lancé une première version bêta de Google News (<http://news.google.com/>) en mars 2002 aux Etats-Unis et en juillet 2003 en France. Ce moteur indexe depuis cette date environ 4 500 sources anglophones et près de 500 sources francophones. En 2002, Ask Jeeves a également fait son entrée sur ce marché avec Ask Jeeves News (<http://news.ask.com/>). Enfin, en 2003, MSN a lancé en version bêta son service MSN Newsbot (<http://fr.newsbot.msn.com/>). Sa technologie Newsbot a depuis été intégrée au moteur MSN dans le monde entier et au site MSNBC.com aux Etats-Unis

Comparaison du nombre de résultats obtenus par MSN, Yahoo! et Google pour le mot-clé "chirac" (versions françaises - le 10 janvier 2005)

	Google	Yahoo!	MSN
Nombre de résultats	4 120	3 208	2 701
Temps de réponse	0,10 secondes	ND	0,09 secondes

Ces moteurs peuvent être interrogés séparément des outils de recherche Web "classiques", en utilisant les onglets spécifiques proposés sur leurs pages d'accueil. Par ailleurs, ils proposent tous une sélection d'actualités organisée par thèmes.

Lorsqu'un terme recherché sur le moteur Web donne des résultats sur leurs sites d'actualités, Yahoo!, Google, MSN et Ask Jeeves intègrent en outre des résultats provenant de leurs outils de recherche d'actualités en tête des résultats de leurs résultats Web.

Résultats provenant d'Ask Jeeves News dans les résultats d'Ask Jeeves



D'autres moteurs - de taille plus modeste - proposent aussi des outils de recherche sur les news. C'est notamment le cas du moteur francophone DeepIndex qui a lancé en 2002 DeepIndex News (<http://www.news.deepindex.com/>), un service présentant l'intérêt d'être centré exclusivement sur la presse francophone. Présent en France depuis octobre 2004, le moteur européen Seekport offre lui aussi un onglet de recherche sur l'actualité (<http://www.seekport.fr/?start=news>).

Dernièrement, Feedster (spécialiste de l'indexation et de la recherche des fils RSS) s'est quant à lui lancé dans l'aventure (<http://news.feedster.com/>) ! Son nouvel outil permet de trier les résultats par pertinence ou par date. Principal intérêt : chaque résultat est agrémenté d'un lien vers le fil RSS de la source d'où il provient.

Services d'agrégation de contenus

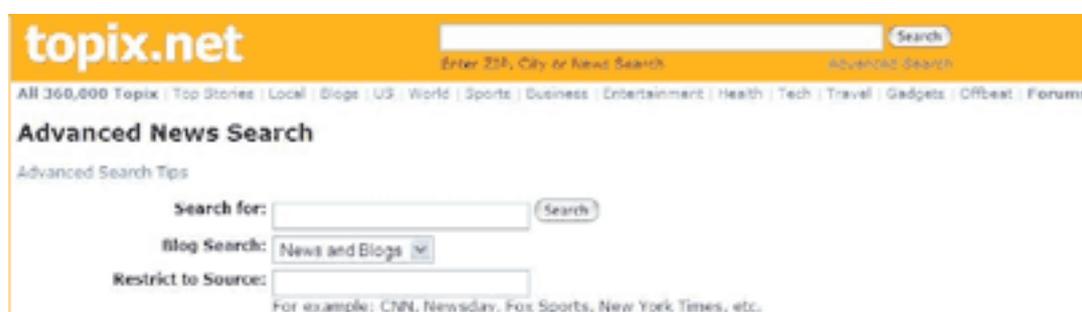
La catégorie des agrégateurs comprend des acteurs qui offrent un outil de recherche aux internautes, et d'autres qui fournissent leurs résultats à d'autres moteurs ou sites Web, sous la forme de fils RSS. Topix.net et Moreover font figure de "leaders" sur ce marché !

Topix.net

Lancé en 2002 aux Etats-Unis par des vétérans de Sun et de l'Open Directory (voir entretien de Chris Tolles), Topix.net (<http://www.topix.net/>) indexe environ 10 000 sources en anglais. Ce service, qui s'autoproclame "plus gros site de news sur Internet", autorise la recherche avancée. Il permet notamment d'exclure certaines catégories de sources (comme les weblogs) ou de restreindre une recherche à un code postal aux Etats-Unis ou au Canada. Il offre également un service d'alertes, ainsi que des fils RSS pour chacune de ses 360 000 catégories. Récemment, le site a ajouté la possibilité pour les membres de poster des commentaires à la suite de chaque article. Il a également ajouté 15 000 weblogs à son corpus de sources surveillées.

La prise de participation de 75% de trois grands journaux américains dans son capital va permettre à la société de poursuivre son développement.

Depuis septembre 2004, Topix.net fournit ses résultats à Ask Jeeves aux Etats-Unis.



Oscoop

Créé en 2004 par Régis Mayo, un développeur français, l'agrégateur Oscoop (<http://www.oscoop.com/>) indexe plus d'une centaine de sources francophones dans 14 catégories, dont l'une est dédiée au référencement de sites Web. Les contenus sont rafraîchis toutes les heures. Oscoop propose également des systèmes d'alertes par e-mail et d'exportation de flux en RSS.

Dernièrement, cet outil a ajouté deux fonctionnalités de clusterisation automatique des contenus et de personnalisation (dans "Mon espace"), avec notamment la possibilité de créer sa propre rubrique.



Trois questions à Régis Mayo (Oscoop)

Comment indexez-vous les sites d'actualités ?

L'intégration se fait manuellement, grâce à notre réseau, on peut intégrer de nouvelles sources d'informations assez régulièrement, elles sont en général au format RSS.

Mais on peut nous contacter, après évaluations, si le site d'information respecte nos critères, il sera intégré dans Oscoop, puis visité par nos robots pour l'indexation des articles.

Quels conseils avez-vous envie de donner aux webmasters ?

Proposer du contenu pertinent régulièrement mis à jour, qu'il respecte une certaine déontologie (pas de racisme, de diffamation, d'insultes). Je pense aux blogs qui peuvent parfois dépasser certaines limites. Techniquement, permettre l'export de



vos infos en RSS, le format pour Oscoop c'est un titre, une description d'une à deux lignes et le lien qui donne sur l'article complet.

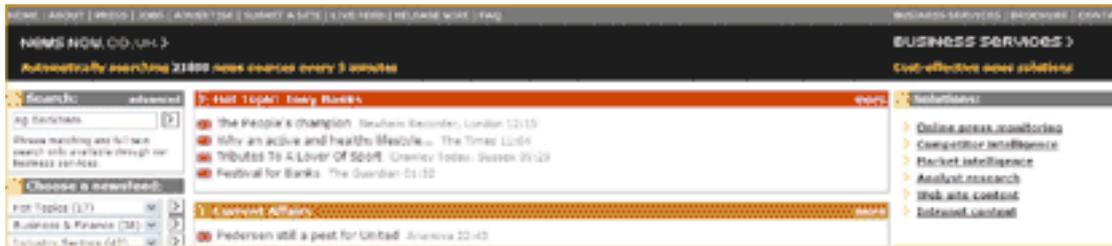
Pouvez-vous nous décrire votre nouveau module de clusterisation ?

Clusterisation, pour les non initiés, peut paraître obscur, en fait c'est un système automatisé qui regroupe les informations du jour par ordre d'importance, et qui a pour but d'extraire les faits marquants du moment. Chaque heure, le calcul est réévalué, afin d'intégrer les nouvelles brèves arrivantes.

NewsNow

Autre outil intéressant, le moteur britannique NewsNow (<http://www.newsnow.co.uk/>) indexe près de 22 000 sources, principalement anglophones, avec des délais de rafraîchissement extrêmement rapides (toutes les 5 minutes). Les contenus indexés sont répartis en 570 catégories.

Pour soumettre leurs fils RSS, NewsNow demande aux webmasters d'intégrer une page spécifique à la racine de leurs sites, comprenant les titres et les URL des articles (voir <http://www.newsnow.co.uk/publishers/specialpage.html>).



Moreover

Moreover est considéré comme l'un des meilleurs agrégateurs de contenus. Après avoir été accessible au grand public pendant plusieurs années, il ne l'est plus, c'est bien dommage ! Propriété de VeriSign depuis 2004, il fournit ses données à différents moteurs de recherche et sites Web. Sa technologie était autrefois utilisée par Altavista. Depuis 2003, elle est également employée par Yahoo. Moreover surveille environ 330 thèmes et indexe plus de 12 000 sources dans 126 pays et 36 langues.



Pour soumettre un weblog sur Moreover, il faut utiliser son serveur de ping ou le formulaire de soumission gratuite (<http://w.moreover.com/cgi-local/site/sourcesubmission.pl>).

Outils spécialisés (moteurs et portails)

FindArticles

Lancé en 2000 par LookSmart, FindArticles (<http://www.findarticles.com/>) indexe plus de 700 publications en anglais. Avantage : il donne accès au texte intégral d'articles publiés depuis 1998. Seul regret, un certain nombre d'articles inclus dans ses archives sont payants. Mais FindArticles n'en reste pas moins un outil extrêmement intéressant pour remettre la main sur un "vieux" article ! Son interface de recherche avancée permet de limiter sa recherche à certains champs du document (titre, corps...), à une source donnée ou à des articles excédant un certain nombre de pages.



Rocketinfo

Arrivé en 2001 sur le marché, RocketNews – désormais appelé Rocketinfo – est un "pure player" de la recherche d'actualités. Son principal outil, Rocketinfo (<http://www.rocketnews.com/>), effectue des recherches dans un panel d'environ 11 000 périodiques et 45 000 blogs. Le groupe édite également des



solutions logicielles payantes de recherche d'actualités pour les entreprises (Rocketinfo Desktop) et un lecteur RSS (Rocket RSS Reader).

Enfin, la société vient de sortir une API en Java (RocketNews Search API) qui permet aux développeurs d'exploiter son service de recherche de news. Il revendique d'ailleurs la place de "premier moteur ayant sorti une API sur un moteur de recherche de news et d'information sur les sociétés".

Daypop

Apparu en 2001, Daypop (<http://www.daypop.com/>) est un outil particulièrement efficace ! Avec une interface "épurée", il indexe près de 60 000 sources (périodiques et weblogs) dans plusieurs langues et autorise la recherche en texte intégral dans les articles. Son spider visite les sites les plus importants toutes les trois heures environ.

Un onglet "Top 40" présente quotidiennement les 40 liens les plus populaires parmi les blogueurs.

NewsLookUp

Lancé en 2004, NewsLookUp

(<http://www.newslookup.com/>), indexe pour sa part un

nombre de sources plus limité que les moteurs précédents. Son intérêt réside principalement dans sa fonctionnalité de classement des résultats par types de sources et par zones géographiques.



180°

Créé tout récemment par deux développeurs californiens, Steve Mallett et Chris Haney, 180° (<http://180n.com/>) est un portail proposant des actualités en anglais *via* une interface simple et conviviale. Chaque article est accompagné d'un "Permalink" (lien permanent) et un lien "Web coverage" pointe, le cas échéant, vers des extraits de weblogs repérés par Technorati ! A tester absolument...

Quatre questions à Chris Haney (180°)

Pouvez vous décrire les caractéristiques de votre outil?

Nous combattons la partialité des médias grâce à la sagesse collective ! Cela signifie que nous inversons l'équation de l'information en permettant à nos lecteurs de décider eux-mêmes de l'importance d'un évènement. Après tout, ne sont-ils pas les mieux placés pour en décider ?

Un internaute peut consulter intégralement "le flot d'informations" livré sur 180° News dans un lecteur RSS pré-configuré. Toutefois, pour ceux qui ne sont pas férus d'informatique, nous réorganisons pour eux, dans un format de lecture très efficace, les informations que nous recevons d'une multitude de sites majeurs d'informations.

Les propriétaires de sites Web ont-ils la possibilité de proposer leurs fils d'infos ?

Il n'y a aucun problème s'il s'agit d'acteurs occupant des niches où le contenu proposé est riche. Il existe très peu de sites d'infos qui occupent des petites niches qui ne soient pas excessivement localisées géographiquement. Les sites d'informations majeurs ont tendance à tous aborder les mêmes sujets, avec seulement un angle un peu différent.

Qu'en est-il des propriétaires de blogs ?

Nous ne nous intéressons pas spécialement à eux en tant que source centrale de fils d'informations. Cependant, nous avons un lien direct pointant vers Technorati qui permet au lecteur de poursuivre sa consultation en lisant les billets postés sur les blogs de la blogosphère. La plupart des blogs sont déjà sur Technorati s'ils ont publié des billets avec des liens pointant vers une actualité publiée sur un site référencé.

De quels financements disposez vous ?

Nous nous autofinançons mais sommes à l'affût de gens intéressés par News 2.0.



Métamoteurs

Les métamoteurs sont eux aussi actifs sur ce marché (voir la lettre R&R du mois de décembre 2005). De même que Mamma et Surf Wax, les trois grands métamoteurs du groupe Infospace

proposent des fonctionnalités de recherche sur l'actualité. La prochaine version du métamoteur français Kartoo devrait aussi inclure une possibilité de recherche sur les news.

Problématiques spécifiques en terme de référencement

Un article concernant l'optimisation du référencement de vos dépêches d'actualité dans les moteurs spécialisés devraient prochainement être disponible dans cette lettre. Voici cependant les grandes lignes de ce qu'il faut savoir dans ce domaine.

Accords de partenariats avec les fournisseurs de contenus

Premier constat : de nombreux outils passent essentiellement par des partenariats avec des fournisseurs de contenus (agences de presse, agrégateurs, et/ou publications) pour leur collecte de données. A l'occasion d'une conférence Search Engine Strategies qui s'est tenue aux Etats-Unis au mois d'octobre, Neil Budde, DG de Yahoo News, a d'ailleurs déclaré : *"Je veux commencer par tordre le cou à une idée assez répandue, suivant laquelle Yahoo News s'appuierait sur une technologie de recherche. En vérité, Yahoo est bâti à partir d'un mélange de données hébergées sur Yahoo. Nous obtenons ces données via des partenariats avec des fournisseurs de contenus, des agences de presse et environ 90 publications avec qui nous travaillons pour amener une partie de leurs contenus sur Yahoo News"* (<http://searchenginewatch.com/searchday/article.php/3556766>).

Référencement des weblogs

De plus en plus d'outils incorporent des blogs à leurs panels de sources surveillées. Depuis octobre 2005, Yahoo US intègre des billets de weblogs aux résultats de Yahoo News. Faites une recherche en ajoutant le mot "blogs" à vos mots-clés et vous verrez apparaître des résultats de weblogs dans un encart séparé à droite des résultats ! Dernièrement, le groupe a conclu un partenariat avec le groupe américain Gawker Media pour publier quotidiennement cinq billets des cinq weblogs qu'il dirige : Gizmodo, Wonkette, Defamer, Lifehacker et Gawker.

Moreover et Topix.net proposent eux aussi des résultats de weblogs. Tout comme Yahoo, ils effectuent un distinguo entre weblogs et périodiques. Jim Pitkow, PDG de Moreover, indiquait d'ailleurs récemment : *"Tandis que les actualités sont davantage régies par des règles de déontologie journalistique, avec des contrôles et des règles, les weblogs affichent surtout des opinions. C'est notre travail de séparer les faits des opinions"* (October 18, 2005 - <http://searchenginewatch.com/searchday/article.php/3556766>).

NB : Pour ceux qui aiment "prendre le pouls" de la blogosphère sur un article, le module de recherche avancée de Technorati (<http://www.technorati.com/>) permet d'effectuer une recherche sur l'URL d'un article. Indiquez votre adresse dans l'encart "Website URL Search" et vérifiez ce qui se dit sur cet article sur les blogs !

Optimiser le référencement des articles

Plusieurs règles doivent être respectées pour garantir le bon positionnement des articles d'un périodique sur les moteurs de recherche d'actualités :

- Tout d'abord, vérifiez vos balises Meta (faites en sorte qu'elles contiennent bien des mots clés ayant trait au contenu de l'article en question) et assurez-vous qu'aucun fichier "robots.txt" n'interdit le passage des spiders sur vos pages !
- N'utilisez jamais une même URL plusieurs fois pour des actualités différentes. En effet, l'emploi d'une même URL rend parfois difficile la détection d'articles nouveaux par les crawlers.
- Pensez à ne publier qu'un seul article par page Web. Les outils qui catégorisent les résultats dans des "clusters" thématiques risquent en effet de rejeter vos pages si elles abordent des thématiques multiples.

- Des titres clairs, factuels et concis facilitent le référencement. En général, soignez également au maximum vos écrits et essayer de publier des articles assez régulièrement. Pour Topix.net, par exemple, la fréquence de publication, le style et la popularité sont des critères de sélection ou de rejet.

- Proposez des fils RSS sur vos pages. Ces fils sont parfois repérés et intégrés, par Yahoo News notamment.

- Enfin, pour un référencement sur Google News, essayez d'inclure au moins trois nombres dans vos URL (par exemple, www.monsite.fr/2006-01-01/actu-referencement). Le moteur britannique NewsNow requiert également l'inclusion de la date (au format inversé) dans le titre du fichier.

Perspectives de développement

Plusieurs critères permettent d'évaluer la performance des outils :

- La fraîcheur d'actualisation de leurs index,
- Le nombre de sources indexées,
- La proportion de dépêches et de communiqués qu'ils incluent dans leurs résultats (la forte proportion de communiqués étant un reproche fréquemment formulé par les internautes),
- Les fonctionnalités de recherche avancée et les fils RSS.

Concernant les "nouvelles" tendances, on constate une adoption de plus en plus massive des fils RSS. Sur sa version américaine, Google News propose depuis peu aux utilisateurs de paramétrer des fils RSS sur des catégories ou des requêtes. Yahoo offre quant à lui déjà cette fonctionnalité sur sa version française.

Autre axe de recherche : la personnalisation. Apparue en juin 2004, Findory (<http://www.findory.com/>) personnalise par exemple les résultats de son moteur sur l'actualité en fonction de ce qu'il a appris des goûts d'un utilisateur membre (au travers des différentes lectures d'articles qu'il a déjà effectuées).

Question aux acteurs des moteurs d'actualité : Comment voyez vous l'avenir de la recherche d'actualités ?

Chris Tolles (Topix.net)

Nous pensons que les gens ont soif d'informations – et que cette soif de sujets d'informations spécifiques qui les intéressent (leur société, leur commune, leurs centres d'intérêt) constitue une part croissante du temps qu'ils passent sur Internet. Il est essentiel d'aider les gens à trouver les informations et les débats, les commentaires les plus pertinents et les plus récents sur les sujets qui les intéressent – le volume d'informations et de commentaires croît si vite qu'il est maintenant "nécessaire" d'avoir un moteur qui vous guide dans cette forêt d'informations.

La principale évolution va consister à inclure les commentaires suscités par les informations dans ce qui était jusqu'alors un fil d'informations unidirectionnel - La transformation majeure qui touche Internet aujourd'hui se trouve dans l'interactivité et dans la communication entre groupes - il n'y a aucune raison pour que cela ne s'applique pas au domaine de l'info.

De plus, l'arrivée de voix alternatives va jouer une part de plus en plus importante dans l'information (voyez l'exemple du tsunami et des attentats de Londres l'an dernier et la part jouée par le photo-journalisme amateur).

L'information est de plus en plus prioritairement consultée en ligne. La capacité qu'a Topix.net de "tagger" automatiquement les articles - qu'il s'agisse de messages postés sur des blogs ou d'informations émanant de sites généralistes – fournit une plateforme de discussion sur tout type de sujets, à chaque ville dans le monde.

2006 promet d'être une année intéressante.

Régis Mayo (Oscop)

Un enjeu majeur, mais plus globalement la recherche d'information aura un rôle encore plus prépondérant dans un avenir proche. On le voit avec le projet européen et son futur moteur de recherche mais aussi avec la Chine, le Japon. L'indépendance de l'information ne peut être

garantie par un seul acteur quel qu'il soit, c'est tout l'enjeu des années à venir.

Chris Haney (180°)

Les algorithmes de recherche bataillent constamment contre les spammeurs. C'est pourquoi je pressens un retour à l'intelligence humaine. Il y aura toujours de la place pour les robots mais ces derniers doivent évoluer plus rapidement que les techniques de spam. Heureusement, les humains s'adaptent plus rapidement que la technique.