

Le référencement dans Google Actualités

[Retour au sommaire de la lettre](#)

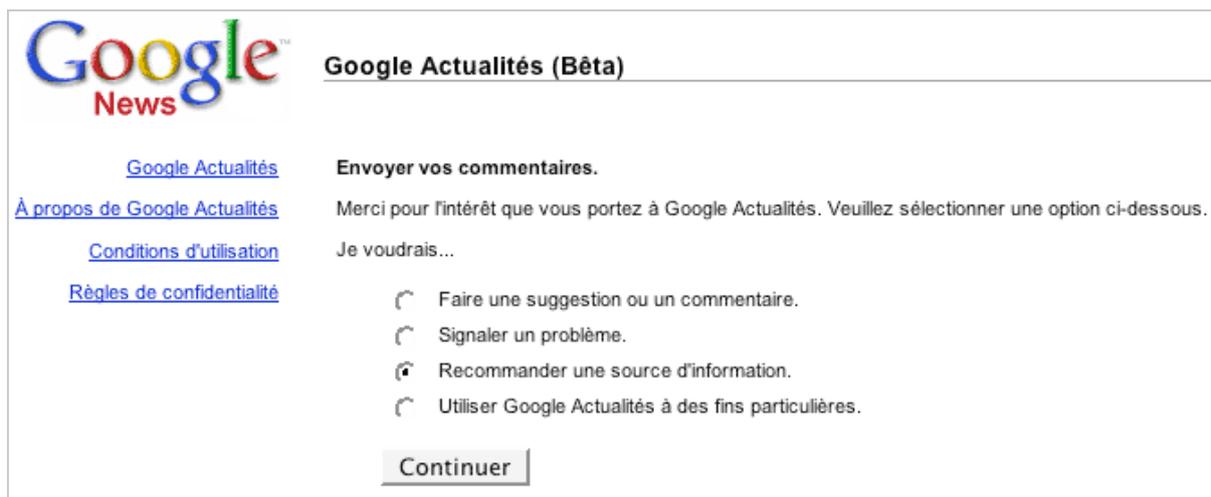
Le moteur de recherche Google Actualités, spécialisé dans la recherche de dépêches et de "news", apporte un trafic important aux sites qui y sont référencés. Quand certaines sources d'informations traînent Google devant les tribunaux pour avoir indexé leur contenu sans autorisation, d'autres tentent désespérément d'y apparaître. Comment voir son site indexé par cet outil ? Comment optimiser ses pages pour obtenir un bon positionnement ou faire s'afficher une image mettant en valeur un lien ? Comment voir son site proposé en page d'accueil de l'outil ? Voici quelques renseignements tirés de notre propre expérience et de nos investigations sur le Web...

A l'heure où des sites d'information belges tentent de faire un procès à Google News pour avoir indexé leurs dépêches d'actualité (<http://actu.abondance.com/2006-38/google-belgique.php>), nous avons choisi ce mois-ci, *a contrario*, de vous parler de la meilleure façon d'intégrer l'index de cet outil. Comment se faire référencer ? Comment optimiser ses articles pour bien apparaître dans les pages de résultats ? Comment voir une image de l'article reprise sur le moteur ? Voici quelques pistes de réflexion pour vous aider à mieux bâtir vos pages d'actualités afin de leur donner une meilleure visibilité sur Google Actualités...

1. Comment se faire référencer sur Google News ?

Dans un premier temps, pour apparaître dans les pages de résultats de l'outil, il faut que Google accepte le référencement de votre site d'information. Si vous êtes un site à forte notoriété, cela ne devrait pas poser trop de problèmes, il y a même de fortes chances pour que cela se fasse sans que vous n'ayiez rien à faire...

Si vous n'avez pas la chance d'être une source d'information à forte notoriété, il vous faut alors demander à être référencé sur le site. Vous devez passer par le formulaire présent à l'adresse <http://www.google.com/support/news?hl=fr> et cocher l'option "**Recommander une source d'information**" :



Google News Google Actualités (Bêta)

[Google Actualités](#)
[À propos de Google Actualités](#)
[Conditions d'utilisation](#)
[Règles de confidentialité](#)

Envoyer vos commentaires.
Merci pour l'intérêt que vous portez à Google Actualités. Veuillez sélectionner une option ci-dessous.

Je voudrais...

- Faire une suggestion ou un commentaire.
- Signaler un problème.
- Recommander une source d'information.
- Utiliser Google Actualités à des fins particulières.

La deuxième étape consiste à donner les indications (URL, nom, adresse e-mail, description) de la source d'informations à indexer :

Google Actualités (Bêta)

Recommander une source d'information.

Pour vérifier si le site que vous avez recommandé est inclus dans Google Actualités, veuillez entrer le nom de domaine dans le champ de texte et cliquer sur le bouton "Vérifier l'inclusion". Bien que nous soyons dans l'incapacité de référencer toutes les sources d'information recommandées, sachez que nous accordons un grand intérêt à vos suggestions.

L'URL de la source :

(exemple : chosun.com)

Google Actualités France BÊTA

Web Images Groupes Actualités plus »

site:abondance.com Recherche Actualités

Résultats 1 - 10 sur un total d'environ 18 pour h:abondance.com. (0,05 secondes)

À la une International France Économie Science/Tech Sports

Microsoft Live Search est officiellement en ligne

Abondance - 11 sep 2006

L'outil Live Search de Microsoft sort de sa phase bêta et est officiellement en ligne depuis aujourd'hui. Le moteur est disponible ...

Nom :

Adresse e-mail :

Objet :

Message :

N'oubliez pas de vérifier (grâce au bouton "Vérifier l'inclusion") que le site n'est pas déjà référencé. En effet, il serait mal accepté par le moteur de soumettre un site qui se trouve déjà dans la base de Google... Vous feriez perdre leur temps aux personnes chargées de vérifier ces données... Lors de la vérification, indiquez l'url sous la forme la plus simple possible (abondance.com, libe.fr, etc.) et surtout sous le domaine où apparaissent les articles mis en ligne (n'indiquez pas www.votresite.com si vos articles sont accessibles sous actu.votresite.com...).

Autre solution : envoyer directement votre demande par mail à l'adresse source-suggestions-fr@google.com...

Sachez simplement que, pour être intégré dans Google News, votre site devra répondre à plusieurs critères :

1. Avoir une **zone "actualités"** ou tout du moins une zone d'infos mise à jour régulièrement. Le site Abondance a, par exemple, trois zones indexées dans Google News :
 - * La zone "actualités" (<http://news.google.fr/news?hl=fr&ned=fr&q=site%3Aactu.abondance.com>).
 - * Son blog (<http://news.google.fr/news?hl=fr&ned=fr&ie=UTF-8&q=site%3Ablog.abondance.com>).
 - * La zone "Infos et question du mois" (<http://news.google.fr/news?hl=fr&ned=fr&ie=UTF-8&q=site%3Adocs.abondance.com>).

Il n'est donc pas nécessaire d'avoir une zone très fortement axée sur l'actualité "fraîche", mise à jour quotidiennement, etc. pour être accepté. Un blog mis à jour chaque semaine peut faire l'affaire et même, comme la zone "question du mois" d'Abondance, une mise à jour mensuelle peut suffire. En revanche, d'une façon ou d'une autre, un être humain, chez Google, va venir évaluer votre site. Evitez, par exemple, dans les jours qui suivent votre demande, les fautes d'orthographe et de frappe sur vos pages... Bref, faites le forçing pendant cette période (mais pas uniquement :-)) pour proposer de l'actu qui "tienne la route" car vous pouvez être sûr que, bientôt, vous allez passer un "examen" à distance..

2. Chaque page d'actualité, chaque information, chaque dépêche doit pouvoir être accédée par l'intermédiaire d'une **url spécifique**. Exemple :

<http://actu.abondance.com/2006-41/exalead.php>

L'actualité comme elle était présentée en 2001 sur le site Abondance, sous la forme d'une seule page par semaine regroupant toutes les dépêches, n'aurait pas été recevable :

<http://actu.abondance.com/actu0141.html>

D'autre part, si un article ne reste pas disponible en ligne pendant les 30 jours pendant lesquels il restera accessible sur Google News, cela peut aussi poser quelques problèmes (cas des articles qui passent en zone d'archives payantes quelques jours après leur publication).

3. Critère assez étrange mais officiellement indiqué par Google : **les urls** de vos pages d'actualité **doivent contenir au moins trois chiffres**.

Exemples :

<http://www.collectifvan.org/article.php?r=4&&id=4632>

http://www.agoravox.fr/article.php3?id_article=14384

http://www.sports.fr/fr/cmcs/cmc/scanner/football/200641/psg-halilhodzic-devra-payer-_109909.html?popup

Vous pouvez vérifier que c'est effectivement le cas de tous les articles référencés sur l'outil... Ces chiffres peuvent désigner la date, le numéro de semaine ou un numéro d'articles, etc. Peu importe. L'essentiel est qu'il y ait ces trois numéros dans l'url...

Ceci est rappelé dans le message que renvoie Google lorsque le site proposé ne propose pas ce type de particularité :

"Merci pour la demande d'ajout sur Google Actualités. Après investigation, nous avons constaté que notre système ne pouvait pas explorer certains de vos articles en raison du format de leurs URL. Pour que vos articles puissent être analysés par le système Google Actualités, leurs URL doivent contenir un nombre composé de trois chiffres au minimum."

Vous savez donc ce qu'il vous reste à faire si cela n'est pas le cas...

4. Certains webmasters ont reçu la réponse suivante suite à leur demande d'inclusion :

"Merci pour votre courrier électronique. Nous avons examiné le site <http://www.votresite.com>, mais nous ne sommes pas en mesure de l'ajouter sur Google Actualités pour l'instant. Nous n'acceptons pas les journaux Web (blogs) sur les actualités ni les sites d'information rédigés et actualisés par des particuliers. De même, nous ne pouvons pas inclure les sites pratiquant la publication ouverte sans processus formel de rédaction. Nous vous remercions d'avoir pris le temps de nous contacter et conserverons votre site afin de l'ajouter si nos modifications nos critères d'intégration. "

Il semble plutôt s'agir ici d'une façon polie de la part de Google d'indiquer que votre site web n'est pas assez "bon" dans son contenu pour être accepté. En effet, on trouve dans Google News de très nombreux blogs rédigés par des particuliers, sans réel processus "formel" de rédaction... Bref, si vous recevez ce type de message, il ne vous reste plus qu'à retravailler la qualité de vos articles et de retenter votre chance d'ici quelques semaines... Et oui, ce n'est jamais très agréable à entendre... Cela dit, si vous "pompez" vos articles un peu partout sur le Web, ne vous étonnez pas de recevoir ce type de réponse...

Sachez, enfin, que le délai d'inclusion dans la base de Google, une fois la demande effectuée, est de un à deux mois... Ne vous impatientez donc pas si, une semaine après votre requête, vous n'êtes toujours pas indexé... Et que Google News "crawle" les pages web de votre site. Lui proposer l'adresse d'un flux XML (RSS, Atom) ne servira à rien, ce n'est pas par ce biais que le moteur indexe votre contenu...

2. Comment assurer une indexation régulière des articles ?

Ce n'est pas parce que votre site est référencé en tant que source d'information sur Google News que tous vos articles, toutes vos dépêches vont être obligatoirement indexés et pris en compte. A

priori (sans que Google ait jamais communiqué de façon officielle à ce sujet), cela semble dépendre de deux critères principaux :

- La taille de la zone éditoriale (l'article en lui-même) proposé : essayez de toujours dépasser les 150 à 200 mots et les 1 200 à 2 000 caractères (espaces compris) pour le corps de l'article, cela devrait fortement augmenter vos chances d'indexation et de prise en compte. Google n'acceptera pas les articles trop courts (moins de 100 mots et / ou moins de 1 000 caractères)... Attention également à la taille du corps de l'article par rapport à la taille totale de la page (charte graphique, zones de navigation, etc.). L'idéal est que ce corps éditorial soit plus important (en termes de taille) que la moitié du contenu total du document... Bref, que l'aspect éditorial soit supérieur à l'aspect "look et navigation"...

- Le nombre d'articles indexés sur un sujet dans l'index du moteur. Si Google estime qu'il a suffisamment d'articles sur un sujet donné qui fait couler beaucoup d'encre (exemple : le récent rachat de YouTube), il se peut qu'il "choisisse" les articles qu'il va indexer pour ne pas crouler sous de trop nombreuses pages. Si le sujet sur lequel vous écrivez un article est très "populaire", tentez de faire un article le plus long possible en termes de mots et de le publier le plus rapidement possible. Cela ne signifie pas non plus qu'il faille faire du "remplissage", n'oubliez pas que vos articles ont pour but d'être lus par des internautes :-)... Une longueur "suffisante" fera peut-être en sorte que votre page sera "retenue" par l'outil mais cela n'est malheureusement pas une garantie...

N'hésitez pas, non plus, à agrémenter votre article d'une illustration (image, photo, graphique, etc.). Il se pourrait bien que cela aide à une meilleure indexation de vos pages (*voir plus loin dans cet article*).

3. Comment apparaître sur la page d'accueil de Google Actualités ?

La page d'accueil du moteur (<http://news.google.fr/>) propose bon nombre d'articles sur des sujets considérés comme "populaires" (certainement le plus souvent cités dans les heures qui viennent de s'écouler), le tout dans plusieurs catégories : "A la une", "International", "France", etc..

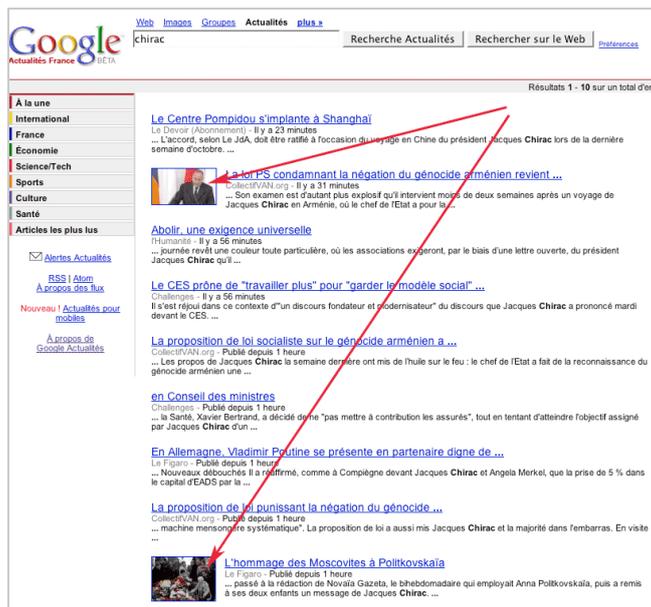
Il semblerait que Google ait mis en place un système de "TrustRank" ou de "NewsRank", ou indice de confiance, fourni en partie par des êtres humains, permettant de donner des priorités à certaines sources d'informations considérées comme plus crédibles. Ainsi, on s'aperçoit rapidement que ce sont souvent les mêmes sources qui apparaissent en "Une" : Le Nouvel Observateur, Libération, Boursier.com, RFI, TSR.ch, Les Echos, L'Express, etc. (voir fin de cet article). Bref, uniquement des sources ayant "pignon sur rue" et contre lesquelles il sera difficile de lutter au niveau de la visibilité, certainement parce qu'elles ont reçu une sorte de "label de qualité" de la part de certains experts chez Google...

Mais il vous sera peut-être possible d'apparaître en page d'accueil sur des sujets plus pointus, sur lesquels les "grands" n'ont pas obligatoirement écrit encore d'articles disponibles en ligne. On s'aperçoit rapidement que sur les "grands titres", il est quasi impossible de "se battre" contre la presse "nationale", mais sur des thématiques plus ténues, vous avez vos chances... A vous, peut-être, d'être rapide et de proposer en ligne un article avant les "grands" pour voir celui-ci repris en une pendant quelques heures... Mais il ne semble pas y avoir de solutions miracles à ce sujet...

Enfin, il est important de savoir un autre point : la quasi totalité de la gestion quotidienne des sites Google Actualités dans le monde est automatisée, sans aucune intervention humaine. Ce ne sont donc pas des éditeurs, contrairement à de nombreux autres sites similaires, qui effectuent des choix d'articles, d'images, etc. Ce sont des algorithmes. Seuls une grosse dizaine de "googlers" travaillent sur l'outil Google News dans le monde... Impossible pour eux d'effectuer un traitement humain quotidien et un quelconque tri des infos qui seraient affichées "à la main"... Seuls le choix et la "notation" des **sources** au départ sont manuels, tout le reste est automatisé...

4. Comment faire apparaître une image ?

Vous vous en êtes certainement rendu compte si vous utilisez souvent l'outil Google Actualités, certains articles sont réhaussés d'une image sur la gauche dans les pages de résultats :



Ce type d'image donne un "focus" indéniable à l'article qu'il souligne. Il y a fort à parier que les taux de clic sur les articles réhaussés par ce type de vignette doit être bien plus fort que lorsqu'il n'y en a pas...

Comment faire, alors, pour faire apparaître ces images ? Il s'agit là d'un mystère pas si simple à percer... Voici cependant quelques indications :

- La présence d'une image ne semble pas avoir de relation avec une quelconque notion de "TrustRank" du site. *A priori*, il ne semble pas que la présence d'une image ait non plus un rapport avec le mot clé saisi sur Google News. Lorsqu'un article donné est souligné d'une vignette dans les pages de résultats, il semble l'être quel que soit le mot clé saisi pour le trouver. Inutile donc d'insérer dans le nom de l'image ou l'option ALT de la balise IMG (*voir ci-après*) certains mots clés dont vous pensez qu'ils peuvent être saisis par les internautes pour trouver l'article en question. Préférez une bonne adéquation entre ces indications et le thème général de l'article, comme nous le verrons par la suite...

- Essayez plutôt d'afficher des images de "grande taille" (plutôt supérieure à 200 pixels sur 200), afin que Google puisse les réduire sous forme de vignette sans trop y perdre en qualité.

- Préférez les formats GIF et JPEG.

- Faites en sorte que le nom de l'image (xxxx.gif ou yyyy.jpg) contienne un ou plusieurs mots explicites et correspondant au contenu de l'article (titre, corps du texte). Exemple : anniversaire-ibm.gif ou jacques-chirac.jpg.

- Remplissez l'option "ALT" de l'image avec le titre de l'article ou sa légende.

Exemple, si le titre de l'article - et de la page - est "Corée du Nord : vers des sanctions à l'ONU", indiquez dans la balise image ces infos :

```
<IMG SRC="Images/coree-du-nord.jpg" alt="Corée du Nord : vers des sanctions à l'ONU" height="265" width="230">
```

Si la légende ou un texte proche de l'image contiennent la phrase "L'essai nucléaire nord-coréen a mis en émoi les grandes puissances mondiales", vous pouvez également indiquer ceci dans votre code HTML :

```
<IMG SRC="Images/coree-du-nord.jpg" alt="L'essai nucléaire nord-coréen a mis en émoi les grandes puissances mondiales" height="265" width="230">
```

- Indiquez la largeur ("width=") et la hauteur ("height=") de l'image dans sa description, comme ci-dessus. Cela aidera éventuellement Google à la réduire sous la forme d'une vignette... A ce sujet, Google semble bien apprécier les images qu'il peut réduire sous un format proche de 60

pixels (hauteur) sur 80 (largeur) ou le contraire pour une image en hauteur (80x60). Tentez peut-être de proposer une image dont la taille correspond à un facteur multiplicateur de ces chiffres...

- Affichez votre image au tout début de votre texte, donc du corps de l'article, juste après le titre. L'alignement peut être à droite, à gauche ou centré, cela ne semble pas poser de problème. En revanche, la présence de l'image au tout début du texte de l'article et après le titre semble essentielle pour qu'elle soit reprise par Google...

- Enfin, il semblerait que l'image ne doive pas être cliquable pour être retenue...

Tous ces conseils devraient vous aider à mieux optimiser la présence d'images extraites de vos articles sous la forme de vignettes dans les pages de résultats de Google Actualités. Il ne s'agit pas, là non plus, de recettes miracles, mais plutôt de "petites astuces" qui devraient améliorer votre situation à ce niveau...

5. Comment mieux positionner un article dans les résultats ?

Peu de surprises *a priori* pour une meilleure optimisation d'une page pour la voir apparaître plutôt en tête de classement sur la saisie d'un mot clé. La plupart des critères pris en compte lors de l'élaboration d'une page web restent valables : bonne optimisation de la balise "TITLE", indication des mots clés importants et descriptifs de l'article en haut de page (dans le titre de l'article - dans le corps de la page - et dans le chapo voire le premier paragraphe), indication du titre dans une balise <H1>, etc.

Bref, rien de bien nouveau à ce niveau-là. Le fait d'indiquer des mots clés dans l'url (www.votresite.com/actu/elections-presidentielles.html) peut jouer également...

Proposez également un couple "balise TITLE / balise meta Description" cohérent l'un par rapport à l'autre et, bien sûr, par rapport au contenu de la page... Mais, là aussi, cela est valable pour le moteur web de Google...

Un point qui nous a paru important dans nos investigations : il nous a semblé que, contrairement au moteur de recherche web, Google News accordait plus d'importance à la présence des mots dans le texte des liens sortants de la page. Si un mot est inclus dans le texte d'un lien, dans un document, cela semble donner un poids plus fort à ce dernier...

Enfin, sachez que Google a déposé un brevet intitulé "[Systems and methods for improving the ranking of news articles](#)" auprès de l'USPTO (l'office des brevets américain) au sujet de l'algorithme de recherche utilisé (ou qui pourrait l'être) dans son moteurs d'actualités. Jean Veronis s'est penché dessus et nous indique, sur son blog (<http://aixtal.blogspot.com/2005/05/google-trustrank-beaucoup-de-bruit.html>) quels sont les 13 "métriques" qui le composent :

1. Le nombre d'articles produit par la source.
2. La longueur moyenne des articles.
3. La "couverture" de la source (combien d'infos ou "stories", la source reprend-elle par rapport à l'ensemble des sources ?).
4. La réactivité de la source ("breaking score"), c'est-à-dire la capacité de la source à diffuser le plus vite possible sur une info donnée.
5. Un indice d'utilisation (en nombre de clics sur cette source).
6. Une opinion humaine sur la source (on retrouve ici la notion de "TrustRank" ou de "NewsRank").
7. Une statistique extérieure d'audience telle que Media Metrix ou Nielsen Netratings, voire la barre d'outils du moteur...
8. La taille de l'équipe, (qui peut-être déterminée par le nombre de journalistes distincts qui signent les nouvelles).
9. Le nombre de bureaux ou agences différentes de la source.
10. Le nombre d'entités nommées originales citées par la source (personnes, organisations, lieux). Ceci sans doute dans l'idée que les sources secondaire enlèvent des infos mais n'en ajoutent que rarement.
11. L'étendue (breadth), c'est-à-dire le nombre du sujets couverts.
12. La diversité internationale.
13. Le style de rédaction, en termes a) d'orthographe b) de grammaire et c) de "niveaux de lecture". Notions pas si simples que cela à évaluer...

Certains de ces critères sont assez complexes à vérifier, il faut bien l'admettre, mais cela donne encore quelques explications et voies de réflexions... Tous ces critères sont certainement pris en compte dans l'algorithme global de pertinence...

6. Comment faire pour ne pas être indexé par Google News ?

Il s'agit ici de la mauvaise nouvelle : il n'existe aucun moyen spécifique de Google News pour voir votre site non indexé par ses robots. En effet, les systèmes de "barrage" proposés par Google pour votre site web sont communs à Google Web ET Google News. Si vous installez un fichier "robots.txt" ou une balise meta "robots" pour barrer l'accès au robots nommés "Googlebot", cela sera valable pour le moteur Web de Google et aussi son moteur d'actualité.

Impossible, en d'autres termes, de barrer l'accès à Google News en laissant la porte ouverte à Google Web, de voir son site indexé sur le moteur web de Google et pas sur le moteur d'actualités. Google lui-même nous a confirmé ce fait. Il est étonnant que le moteur leader ne propose pas ce type de possibilité à l'heure actuelle et c'est peut-être là la source de certains de ses ennuis juridiques à ce niveau... Même s'il y a une certaine logique à ce comportement. En effet, on voit mal pourquoi un site accepterait l'indexation de ses dépêches d'actualités sur le moteur Web et pas sur le moteur de news... Ceci dit, le choix pourrait quand même être laissé à l'éditeur du site, ce qui n'est pas le cas à l'heure actuelle...

Si vous désirez donc voir votre site éliminé des résultats sur les deux outils de recherche, vous pouvez mettre en place un fichier "robots.txt" contenant les lignes suivantes :

```
User-agent: Googlebot  
Disallow: /
```

Ou utiliser les deux balises meta "robots" suivantes au choix :

```
<META NAME="ROBOTS" CONTENT="NOINDEX, NOFOLLOW">  
<META NAME="GOOGLEBOT" CONTENT="NOINDEX, NOFOLLOW">
```

Plus d'infos ici à ce sujet déjà traité dans des articles précédents de cette lettre :

<http://www.google.com/support/webmasters/bin/answer.py?answer=35302>

<http://www.google.com/support/webmasters/bin/answer.py?answer=35303>

7. Conclusion

L'indexation d'une source d'information dans le moteur de recherche d'actualité de Google est une source évidente de trafic. Encore faut-il que votre site soit accepté (dans ce cas, ce sont des êtres humains qui prendront la décision chez Google) et que vos articles soient bien indexés et optimisés pour apparaître sur la page d'accueil et / ou les pages de résultats et là, ce sont les processus d'automatisation mis en place par Google qu'il vous faudra "apprivoiser"... La lecture de cet article devrait vous donner quelques pistes afin de tenir d'un certain nombre de critères importants qu'il vous faudra prendre en compte lors de l'écriture de vos prochaines dépêches...

Il existait déjà des formations apprenant à "écrire pour le Web", peut-être bientôt verrons-nous des méthodologies d'apprentissages d'articles pour les moteurs de recherche d'actualité...

PS : à noter un site qui récupère et analyse la liste des sources utilisées par Google Actualités (avec un "top 25" par mois, par an, etc), cette liste n'étant pas officiellement proposée sur son site par le moteur... Bref un site de statistiques très utile :

<http://www.privateradio.org/blog/i/google-news/reports/fr/index.php>

Voici par exemple le "Top 10" des sources d'information le plus présentes en 2006 sur la page d'accueil du moteur :

1. Nouvel Observateur
2. Le Figaro
3. L'Express
4. Libération
5. Boursier.com
6. Le Monde
7. TF1

8. Reuters.fr
9. France
10. La Tribune.fr

Difficile, on le voit, pour un "petit" site de "faire son trou" en *homepage* de Google Actualités...

Enfin, pour clôturer le sujet, quelques articles au sujet de Google News et du référencement de sites web dans son index :

Un article du site News Scientist sur le TrustRank :

<http://www.newscientist.com/article.ns?id=mg18624975.900>

Des travaux de l'université de Stanford sur ce même sujet ("Combating Web Spam with TrustRank" ou la notion de "confiance" mise en équations) :

<http://dbpubs.stanford.edu:8090/pub/2004-17>

Un webmaster qui explique comment il a spammé Google News :

<http://3couleurs.blogspot.com/2005/10/jai-spamm-google-news.html>

A lire si vous avez envie d'être blacklisté :-)