

Optimiser son site et son contenu pour Google Actualités

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

L'outil de recherche Google Actualités ("Google News") est devenu un incontournable du Web, à tel point qu'il génère sur certains sites spécialisés près de 70% du trafic "moteurs". Comment optimiser son site pour Google News ? Comment voir sa source d'information acceptée sur l'outil ? Une spécialiste du domaine, responsable du référencement pour un grand groupe de presse, tente ici de répondre de la façon la plus précise possible à toutes ces questions ...

L'indexation d'un site dans Google Actualités offre des perspectives de trafic gratuit extrêmement intéressantes, en particulier pour les sites proposant de l'actualité fraîche et variée. Sur les sites de news majeurs, le pourcentage du trafic moteur en provenance de Google Actualités peut aller jusqu'à **70% du trafic moteur global** !

Pour être présent dans l'index de Google Actualités, il est impératif de respecter certaines contraintes techniques et éditoriales de construction des pages. Pour être bien positionné dans Google Actualités, il faut y adjoindre certaines contraintes éditoriales. Nous allons les décrire dans cet article...

Respecter ces critères est particulièrement stratégique pour les contenus dont l'attrait à une durée de vie inférieure ou égale à 30 jours (ou contenu « chaud »). Mais d'expérience, on se rend compte que Google Actualités indexe parfois aussi du contenu plus « froid ». Il peut donc être intéressant de respecter ces critères de manière plus générale dans un site et pas seulement pour la/les sections actu.

Pour être référencé sur Google Actualités, il faut simplement en faire la demande à Google via un formulaire dédié. Mais avant de faire cette demande, il faut vérifier que le site/blog a soumettre est éditorialement et techniquement conforme au cahier des charges de Google Actualités : un site qui ne le serait pas pourra être accepté (si la qualité de son contenu le justifie) mais ne sera pas ou peu crawlé ; ne générant ainsi pas de trafic.

A. Les critères de base

- Chaque article doit disposer de sa page web dédiée

Proscrire les pages listant plusieurs "brèves", les listes d'articles ne pointant pas vers la page dédiée à chaque news et les pages contenant le contenu intégral de plusieurs articles. Ces types d'articles seront systématiquement refusés par Google Actualités.

- Chaque article doit disposer d'un identifiant unique présent dans son url

L'url doit contenir un ID unique composé d'au moins trois chiffres (Attention : le nombre ne doit pas être composé d'un ID à quatre chiffres ressemblant à une année). Il n'est pas obligatoire que cet ID soit présent en fin d'url, mais il doit être présent. Les urls ne contenant pas d'ID ne sont tout simplement pas crawlées.

Exemples d'urls à id conformes pour Google Actualités :

www.monsite.com/actu/1234/titre-article.html
www.monsite.com.actu/titre-article-1234.html

Google avait communiqué en février 2009 sur le fait que ce critère n'était plus obligatoire (l'info viendrait, semble-t-il, de Google UK). Si cela n'est plus le cas aux USA ou dans le monde anglophone (ce qui reste à confirmer), il n'en est rien en France :

toute page web ne disposant pas d'un ID dans son url est toujours ignorée par le *spider* de Google Actualités.

- L'article doit être en ligne au moins 30 jours

C'est la durée d'indexation des articles dans Google Actualités. La commande "site:" dans Google Actualités permet de savoir combien d'articles d'un site sont indexés. Au-delà de 30 jours, les articles ne remontent plus dans les pages de résultat.

- Les images

Les images (et en particulier l'image d'illustration principale de l'article) doivent être non cliquables. Il existe cependant des techniques de contournement pour disposer d'un zoom sur les images sans passer par un lien classique ` ...`

Pour disposer d'une vignette à côté de son article dans les pages de Google Actualités, il faut respecter un format qui permette à Google de la générer sans recadrage. Pour cette raison, il faut afficher un visuel d'illustration principal de l'article dont la taille correspond à un facteur multiplicateur de 60x80 en mode portrait, ou 80x60 en mode paysage. Les images doivent être en gif ou jpeg uniquement et les attributs height et width de la balise img doivent être remplis dans le code (au moins la largeur OU la hauteur). Il en est de même pour l'attribut alt (ou description courte de l'image). Les articles illustrés d'une image en PNG ne permettront pas la génération d'une vignette dans les pages de résultat de Google Actualités.

Le non respect des *guidelines* images n'entraînera pas une non indexation des articles. En revanche, il sera alors impossible que les articles soient illustrés d'une vignette.

- Balisage sémantique

Un balisage sémantique correct est beaucoup plus important pour le positionnement dans Google Actualités que pour Google "classique".
Le titre doit être contenu dans une balise `<h1>`, le chapô (optionnel) dans une balise `<h2>`

- Le contenu textuel de l'article

S'il n'existe pas de longueur idéale pour être bien positionné sur Google Actualités, il existe en revanche une longueur minimale pour être indexé.
Les articles dont le texte n'atteint pas 1200 caractères (espace et chapô compris) seront la plupart du temps rejetés (mais parfois indexés, sans qu'il semble y avoir une raison pour que l'indexation soit effective dans un cas et pas dans un autre...). Il est possible que Google Actualités soit plus souple sur ce critère avec les sources disposant d'un fort ranking score (voir ci-dessous).
Attention, il s'agit du texte de l'article "nu", sans prise en compte du reste du contenu textuel de la page et des blocs de rebond.
En général, Google se réfère à la balise `</h2>` (ou, à défaut `</h1>`) pour identifier le début de l'article et à la date ou à la signature de l'article pour en identifier la fin.

La présence d'un chapô n'est pas obligatoire mais cependant recommandée par Google. S'il est présent, il ne doit pas être cliquable.

Il est recommandé par Google :

- d'afficher la date et l'heure de publication de l'article (ceux-ci ne doivent pas être cliquables)
- d'afficher le nom de la rubrique initiale de publication de l'article, ou les thèmes associés (ceux-ci ne doivent pas être cliquables)

- Le titre de l'article

Il doit être court, plus court que la longueur maximale autorisée pour Google « classique » : 60 caractères au maximum selon les *guidelines* de Google. Dans la pratique, un titre « passe » souvent jusqu'à 70 caractères espaces compris. En cas de

dépassement, le titre sera tronqué dans les pages de Google Actualités, mais cela n'entraînera pas la non indexation de l'article.

Comme pour Google classique, le titre est un élément très important pour le positionnement. Sur Google Actualités, il a même un poids encore plus important. L'équation a résoudre pour bien choisir son titre est celle de la part de l'optimisation « SEO » versus celle du *teasing* purement éditorial. En effet, du CTR est dépendant le classement de l'article, et donc sa remontée éventuelle en homepage. Pour trouver le bon mix, il faut, selon son site/blog et sa cible tester, tester, tester, et encore tester...

- La densité de mots clés

Celle-ci ne semble avoir aucune incidence ni sur l'indexation, ni sur le positionnement.

Google donne un certain nombre d'informations sur les contraintes techniques liées à Google Actualités sur son aide en ligne dédiée :

http://www.google.com/support/news_pub/bin/topic.py?hl=fr&topic=11665

- Le Sitemap News

L'expérience montre qu'il améliore grandement le taux d'inclusion des articles : Google n'a plus à trouver lui-même sur le site quels sont les nouveaux contenus d'actualités à indexer : ils lui sont fournis directement à la publication.

Il est possible de soumettre le Sitemap news comme n'importe quel Sitemap « classique » via les Google Webmaster Tools. Il est souhaitable également d'ajouter son url dans le Sitemap index du site.

Ne doivent être incluses dans le Sitemap que des urls d'articles de type « actu », dont la date de publication est inférieure à 30 jours. On peut éventuellement y adjoindre des pages de listes d'actu. La mise à jour du fichier doit se faire si possible en temps réel, à la création d'un nouvel article.

Le Sitemap ne doit pas contenir plus de 1 000 urls (si besoin, il faudra créer plusieurs Sitemaps news).

L'urlset diffère de celui du Sitemap « classique » :

```
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
xmlns:news="http://www.google.com/schemas/sitemap-news/0.9">
```

Les balises du Sitemap News sont les suivantes :

Balise		Usage
<loc>	<i>obligatoire</i>	URL de la page
<news:news>	<i>Optionnelle mais conseillée</i>	Début des META Données propres à Google News qualifiant l'article.
<news:publication_date>	<i>Optionnelle mais fortement conseillée</i>	Date de publication de l'article au format W3C (YYYY-MM-DDThh:mmTZD).
<news:keywords>	<i>optionnelle</i>	Contient une liste de mots clés séparés par des virgules, dont l'objectif est de guider Google dans la classification des contenus. Des catégories existent dans Google Actualités http://www.google.com/support/webmasters/bin/answer.py?answer=42993 mais il reste très simple d'utiliser les mots clés associés au contenu par le rédacteur, sans utiliser la catégorisation générale proposée par Google News.
<news:stock_tickers>	<i>optionnelle</i>	A utiliser pour les articles à contenu financier : liste des valeurs, fonds d'investissement, etc. séparés par une

	virgule. Chaque valeur peut également être précédée du nom de la bourse d'où provient sa cotation, par exemple, "NYSE:HD"
--	---

L'ensemble des recommandations Google relatives aux Sitemap news sont disponibles en ligne sur l'aide aux éditeurs :

http://www.google.com/support/news_pub/bin/topic.py?hl=fr&topic=11666

B. Une gestion optimisée du taux de refus

Le premier élément à prendre en considération est que Google ne repasse jamais sur un article : il faut qu'on qu'il soit optimal dès la publication. S'il est rejeté, c'est fini.

Pour générer le maximum de trafic possible, il est donc primordial de faire baisser le taux de refus !

Google communique tous les 15 jours une liste des articles refusés par Google Actualités dans le Google Webmaster Tools (dans la section *Diagnostic/Erreurs d'exploration/Actualités*).

Pour chaque article refusé, Google donne un motif. Mais leur interprétation n'est pas toujours aisée...

La liste est accessible (mais curieusement difficile à trouver...) sur l'Aide aux éditeurs à cette adresse: <http://www.google.com/support/webmasters/bin/answer.py?hl=fr&answer=93994>

La voici également ici avec nos commentaires :

Article trop court	Le contenu de l'article fait moins de 1200 signes
Article trop court par rapport au texte de la page	Il y a trop de blocs de remontée de contenu sur la page par rapport au texte de l'article lui-même. Il faut dans ce cas augmenter la longueur de l'article, ou supprimer des blocs de remontée.
Article fragmenté	Le spider de Google news semble détester les successions de paragraphes courts, ou les mises en page un peu « exotiques ».
Article trop long	C'est une erreur que l'on ne rencontre que rarement, et qui est le plus souvent causée par l'absence de signature ou de date en bas de l'article, ce qui empêche le spider de différencier ce qui est du contenu de l'article et ce qui ne l'est pas.
Date introuvable	On ne rencontre plus cette erreur si la date de publication de l'article est présente dans le Sitemap news et/ou dans l'article.
Date trop éloignée	Si Google news n'a pas crawlé l'article dans les 3 jours suivant sa publication, il choisit de ne pas l'indexer, estimant qu'il n'est plus assez « frais ». Pour éviter ce type de refus, le Sitemap news est utile mais pas suffisant : il faut également augmenter la production éditoriale pour pousser le spider à revenir plus souvent. Le problème peut venir également d'un éparpillement des contenus de type news à l'intérieur du site/blog, ce qui rend la tâche du spider plus ardue.
Article vide	Nous n'avons personnellement jamais rencontré ce type de refus et il est difficile à évaluer précisément : serveur web trop lent au moment du crawl ? Bug du site ou bug du spider de Google Actualités ?
Échec de l'extraction	Nous n'avons personnellement jamais rencontré ce type de refus. Il provient très certainement d'une construction fonctionnelle ou technique non conforme du <i>template</i> de la page article.
Aucun lien trouvé	Nous n'avons personnellement jamais rencontré ce type de refus. Mais il est vraisemblable qu'il concerne les pages « orphelines »

	(aucun backlink).
Aucune phrase	C'est de tous les motifs de refus l'un des plus fréquents et l'un des plus frustrants car il apparaît de manière erratique. Serveur web trop lent au moment du crawl ? Bug du site ou bug du spider de Google Actualités ? Dans tous les cas, le problème n'est ni dû à la construction du template ni au contenu éditorial.
Balise "noindex" détectée	"La page HTML de l'article contient une balise Meta "noindex" qui empêche Google d'indexer la page. Essayez de supprimer la balise Meta "noindex"."
Redirection hors du site	"La section ou la page d'article redirige vers une URL appartenant à un autre domaine. Les pages de section et les articles doivent tous faire partie du domaine du site indiqué dans Google Actualités."
Page trop longue	Il ne s'agit pas d'un nombre de signes mais plutôt d'un poids maximum autorisé pour le HTML du contenu article : "Les pages ne peuvent pas excéder 256 Ko". Il est vraisemblable qu'il s'agisse ici d'une limite technique imposée par Google pour ne pas surcharger ses serveurs.
Titre non autorisé	Ce type d'erreur résulte souvent d'une non reprise dans la balise <title> du titre contenu dans la balise <H1> : ou d'une absence de cette dernière.
Titre introuvable	Les causes sont multiples pour ce type de refus : titre trop court, titre trop long, non correspondance entre le contenu de la balise <title> et la balise <H1>, ...
Échec de décompression	"Googlebot a constaté que cette page était compressée, mais n'est pas parvenu à la décompresser."
Type de contenu non pris en charge	"La page contient une variable content-type HTTP. Ce type de variable n'est pas pris en charge par Google Actualités. La variable content-type des articles doit être text/html ou text/plain."

C. L'optimisation avancée

Une fois levés tous les freins techniques et éditoriaux, l'objectif est maintenant de générer beaucoup de trafic, ou en tout cas d'en générer plus que ses concurrents directs.

Or, le classement dans Google Actualités semble de prime abord opaque, voire aléatoire ou injuste. En réalité, il obéit à des règles très précises. Contrairement à une idée reçue, il n'y a pas de manipulation manuelle des classements de la part de Google, hormis lors de la classification initiale des sources qui n'est pas réalisée par une machine.

Il faut savoir que tous les sites soumis à Google Actualités ne sont pas acceptés. Il faut respecter un certain nombre de pré-requis qui sont vérifiés manuellement par les équipes de Google Actualités. Il faut par exemple disposer d'une équipe éditoriale (même réduite), que les articles ne soient pas à visée commerciale et être propriétaire du contenu, etc.

Il semble exister un *ranking* des sources sur Google Actualités, dont le plus haut niveau pourrait être qualifié de "Premium". Ce statut semble pouvoir s'acquérir ou être obtenu directement à l'inclusion (il serait dans ce cas donné par les équipes de Google). Voir cette URL à ce sujet : <http://aixtal.blogspot.com/2005/05/google-trustrank-beaucoup-de-bruit.html>.

Quels sites sont susceptibles d'être automatiquement qualifiés de "Premium" ? Il s'agit des principaux organes de presse papier dont la qualité est le sérieux n'est plus à prouver et qui sont des « sources de confiance », par exemple Le Monde, Le Figaro, Libération, L'express, etc.

Comment augmenter son "Source Ranking", voire devenir "Premium" ?

Il s'agit ici de suppositions basées sur l'analyse et l'expérience, à l'exclusion de toute communication officielle de la part de Google... Voici quelques pistes de réflexion :

- produire beaucoup d'articles chaque mois, de manière régulière sur la journée, la semaine et le mois ;
- produire un contenu riche et varié ;
- produire un contenu en adéquation avec des *trending topics* du moment (donc correspondant aux recherches des internautes : les sujets de niche semblent avoir beaucoup de mal à émerger dans Google Actualités).
- sortir des scoops.
- voir ses articles globalement très cliqués par les internautes.

Les notions-clés du ranking dans Google Actualités semblent donc être : scoops, CTR très élevé, variété et quantité de contenus produits.

Lorsqu'un sujet est très recherché et que de nombreux articles sont produits dans un laps de temps très court par des sites différents, Google va d'abord privilégier le premier article publié (sauf s'il vient d'une source disposant d'un ranking faible) puis accorder la première place à l'article le plus cliqué. Il y a d'ailleurs une forte corrélation entre les articles les plus lus et ceux qui remontent en homepage.

Au bout d'un moment qui varie en fonction du sujet et du nombre d'articles produits, Google accordera plutôt la première place à l'article le plus frais ou apportant des informations complémentaires par rapport aux articles précédents.

=> Il existe donc un *turn over* des critères en fonction du cycle de vie de l'information.

Google semble également parfois privilégier les sources locales lorsque l'information contenue dans l'article a un intérêt local, ou si elle a un intérêt national mais un ancrage local.

D. Soumettre / Contacter / Déréférencer son site de Google Actualités

Il est très simple de soumettre son site à Google Actualités. Il suffit de remplir un formulaire : http://www.google.com/support/news_pub/bin/request.py?contact_type=suggest_content&hl=fr

Un mail automatique est ensuite envoyé, auquel il faut répondre (visiblement une procédure anti spam). Le délai moyen entre la demande d'inclusion et le début de l'indexation est de 3 semaines environ. Mais certains sites ont été ajoutés par le passé dans Google News directement par les équipes de l'outil, sans qu'aucune demande préalable n'ait été effectuée...

Formulaire de contact Google News

Une fois son site inclus, il est parfois utile de contacter Google. Il existe des cas où des sites ne sont plus crawlés du jour au lendemain, ou ne le sont plus que partiellement. Le plus souvent dans ce genre de cas, il semble s'agir de bugs.

Dans ce cas, l'expérience a montré qu'il n'y a rien de bon à attendre que cela s'arrange tout seul : c'est rarement le cas. Il faut contacter directement les équipes de Google pour leur signaler le problème. En général, pour les cas simples, tout rentre dans l'ordre sous quelques jours. Pour les cas plus complexes, cela peut parfois être beaucoup plus compliqué...

Pour contacter les équipes de Google Actualités, utiliser ce formulaire : http://www.google.com/support/news_pub/bin/request.py

Informer Google Actualités de changements sur son site/blog

Il n'est pas possible de soumettre un article de manière unitaire. En revanche, en cas de modifications sur le site non prises en compte (ajout d'une nouvelle section, changement de nom de domaine, etc.), il est possible de le signaler ici :

http://www.google.com/support/news_pub/bin/topic.py?hl=fr&topic=11708

Référencer sa chaine Google YouTube sous Google Actualités

Il est à présent possible de soumettre à Google News sa chaine vidéo Youtube afin de voir ses vidéos postées sur Youtube remonter dans Google Actualités.

Les informations détaillées sont disponibles sur les sites de Google et de Youtube sur ces URL:

http://www.google.com/support/news_pub/bin/answer.py?answer=93985

<http://www.google.com/support/youtube/bin/answer.py?answer=141804>

A noter : à ce jour, Youtube est la seule plateforme vidéo permettant le référencement des vidéos de votre site dans Google Actualités.

Déréférencer un article en particulier ou la totalité de son site/blog de Google Actualités

Il est possible de déréférencer tout ou partie de son site de Google Actualités *via* des formulaires de contact dédiés sur l'aide éditeurs Google Actualités ici :

http://www.google.com/support/news_pub/bin/topic.py?hl=fr&topic=11709

E. Conclusion :

On l'a vu, même si être inclus dans Google Actualités est relativement simple, les contraintes techniques et éditoriales pour être indexé sont nombreuses.

Au delà des bonnes pratiques énumérées ci-dessus, il faut un certain temps d'expérimentation et d'habitude avant de commencer à trouver les réglages éditoriaux optimums car l'expérience montre que ceux-ci changent en fonction de la thématique et du nombre de sites en concurrence.

Un mot d'ordre : tester, tester, tester !

Le trafic Google Actualités augmente au fur et à mesure du temps et de l'expérience acquise, de manière en général relativement linéaire (cela est dû à l'importance du taux de clic et du *source ranking* dans les classements). Maintenant, à vos claviers pour tenter l'expérience !

Virginie Clève, responsable SEO chez Prisma Presse.

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

<http://abonnes.abondance.com/blogpro/2009/10/octobre-2009-optimiser-son-site-et-son.html>