

Comment bien gérer les 404 dans une optique SEO ?

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Tout site Internet génère des erreurs 404 ("Page not found"). Ces pages peuvent-elles réellement nuire au référencement naturel, et si oui comment faire pour les traiter efficacement ? Nous verrons dans cet article pourquoi cela peut parfois nuire à votre positionnement ou à l'indexation de vos contenus. Mais surtout, nous allons traiter les différentes méthodes pour en récupérer la liste complète et les différentes façons de les corriger de manière optimale pour votre SEO.

Les erreurs 404

En général, une erreur 404 signifie que votre page renvoie le code HTTP « 404 Not Found » en réponse au chargement d'une URL précise. Pour obtenir la liste complète de ces codes d'en-tête HTTP, consultez cette page :

http://fr.wikipedia.org/wiki/Liste_des_codes_HTTP

Avoir des pages 404 est une chose courante et normale pour un site Internet, et elles peuvent provenir de plusieurs causes :

- Un fichier ou une page est temporairement indisponible ;
- Vous avez modifié l'URL d'un contenu ;
- Vous avez modifié le nom d'un fichier (image, vidéo...) ;
- Un de vos scripts (PHP par exemple) provoque des bugs et des liens erronés ;
- Un autre site (ou une page de votre site) a fait un lien erroné vers le vôtre ;
- Etc.

En soi, les erreurs 404 ne nuisent pas à votre référencement naturel puisque Google et les autres moteurs de recherche savent parfaitement gérer ce type de code HTTP.

Les problèmes SEO des erreurs 404

Tant que cela reste ponctuel, l'impact sur votre référencement naturel sera donc imperceptible. Mais cela peut poser deux types de problèmes.

L'indexation

Le premier problème est celui du frein à l'indexation. En général, plus un moteur de recherche va trouver de liens qui renvoient des erreurs 404 ou des redirections, plus il aura tendance à trouver moins pertinent l'ensemble du site.

Il va donc en général ralentir son indexation pour vos "vrais" contenus car il devra faire le tri constamment entre vos publications réelles et celles qui sont en erreur ou qui ont été déplacées. Il aura d'ailleurs le même comportement lorsqu'il trouve du contenu dupliqué sur un site.

Heureusement, cette problématique ne se rencontre que si le nombre d'erreurs 404 est important. Si ce n'est pas le cas, cela n'aura pas d'impact réellement visible.

La perte de popularité

Ce second problème est sans doute le plus gênant. Vous avez des erreurs 404 sur des URL qui ont parfois un nombre de liens important depuis d'autres sites ou depuis vos propres contenus.

En termes de maillage interne, cela va réduire l'efficacité de la structure de votre site. Si vous mettez en effet en place un maillage logique et efficace, celui-ci va perdre de son efficacité si trop de vos contenus renvoient une erreur 404.

Pour les contenus qui ont des liens depuis des sites externes, ils reçoivent une popularité plus ou moins importante qui sera alors perdue à cause du code HTTP d'erreur. Vous perdez donc du poids aux yeux de Google et cela vous fait donc potentiellement perdre du positionnement. En clair, vous perdez des backlinks, pourtant si difficiles à obtenir en temps normal !

L'erreur 404 a donc un impact indirect sur le référencement d'un site, qui peut parfois être important.

Comment les connaître ?

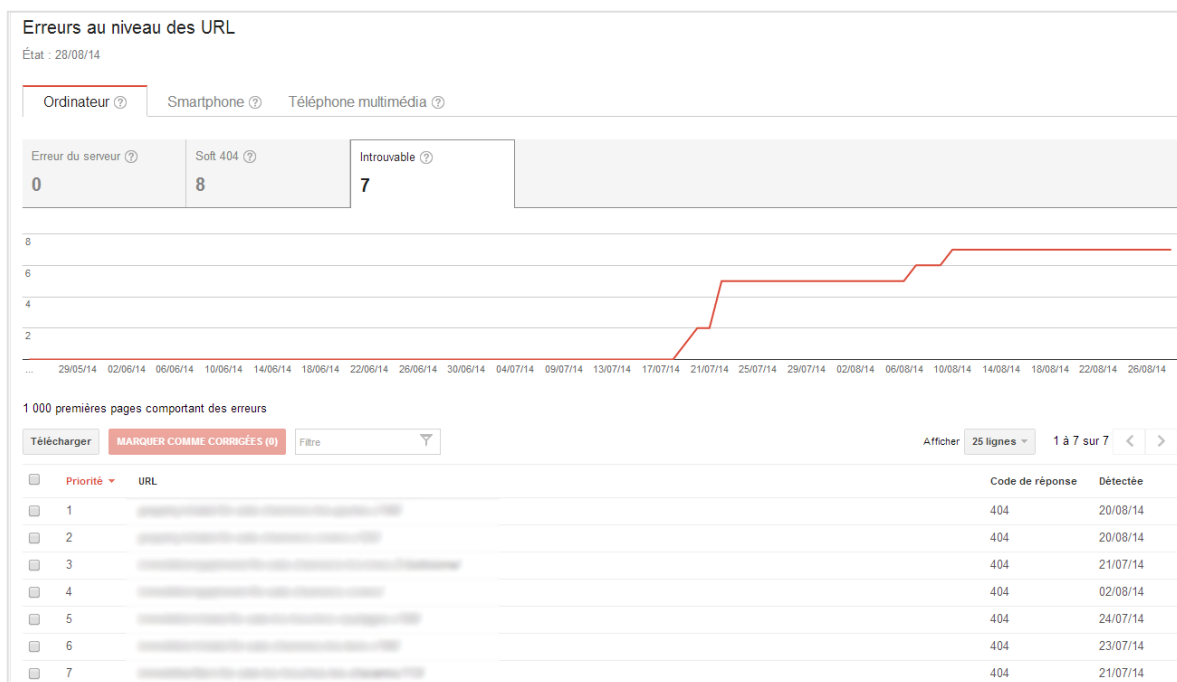
C'est la première étape : avant de les traiter, encore faut-il en avoir connaissance.

Le centre Webmaster de Google

Le centre pour Webmaster de Google vous fournit la liste des pages rencontrées par le robot du moteur de recherche et qui posent problème sur votre site Internet. On y retrouve :

- Les pages 404 ;
- Les soft 404 (nous y reviendrons plus tard) ;
- Les redirections ;
- Les erreurs serveurs.

Dans le menu « Erreurs d'exploration », vous aurez donc la liste des pages Introuvables de votre site.



Sachez que chaque erreur est cliquable dans le tableau, et vous donne des informations très intéressantes, notamment :

- La dernière date où l'erreur a été rencontrée ;
- La date à laquelle l'erreur a été pour la première fois rencontrée ;

- Un onglet « Dans les Sitemaps » si l'erreur provient d'un fichier sitemap.xml ;
- Un onglet « Référencée sur » si l'erreur provient d'un mauvais lien en interne ou en externe.

Introuvable

URL : [http://www.abondance.com/.../.../...](#)

Informations sur l'erreur Dans les sitemaps Référencée sur

Dernière exploration : 20/08/14
Première détection : 09/08/14

Impossible d'explorer cette URL, car elle redirige vers une page inexistante. En général, les erreurs 404 n'ont aucune incidence négative sur les performances de votre site dans les résultats de recherche, mais elles peuvent vous aider à en améliorer l'expérience utilisateur. [En savoir plus](#)

[Marquer comme corrigées](#) [Annuler](#) [Explorer comme Google](#) ?

Attention, le bouton « Marquer comme corrigées » ne signifie pas que Google va considérer la page d'erreur comme Résolue : cela l'enlèvera uniquement de votre tableau. Il faudra attendre que le robot de Google revérifie votre contenu pour que votre correction soit prise en compte (si tel est le cas, bien sûr...).

Xenu, Integrity et autre logiciels de crawl

Il vous est aussi possible de lancer un scan de votre site à l'aide d'un logiciel de crawl. On peut notamment citer :

- Xenu Link Sleuth <http://home.snafu.de/tilman/xenulink.html> ;
- Integrity <http://peacockmedia.co.uk/integrity/> ;
- Yooda SeeUrank Falcon <http://www.yooda.com/> ;
- ...

A vous ensuite de trier les données de ces scans pour ne faire ressortir que les erreurs 404.

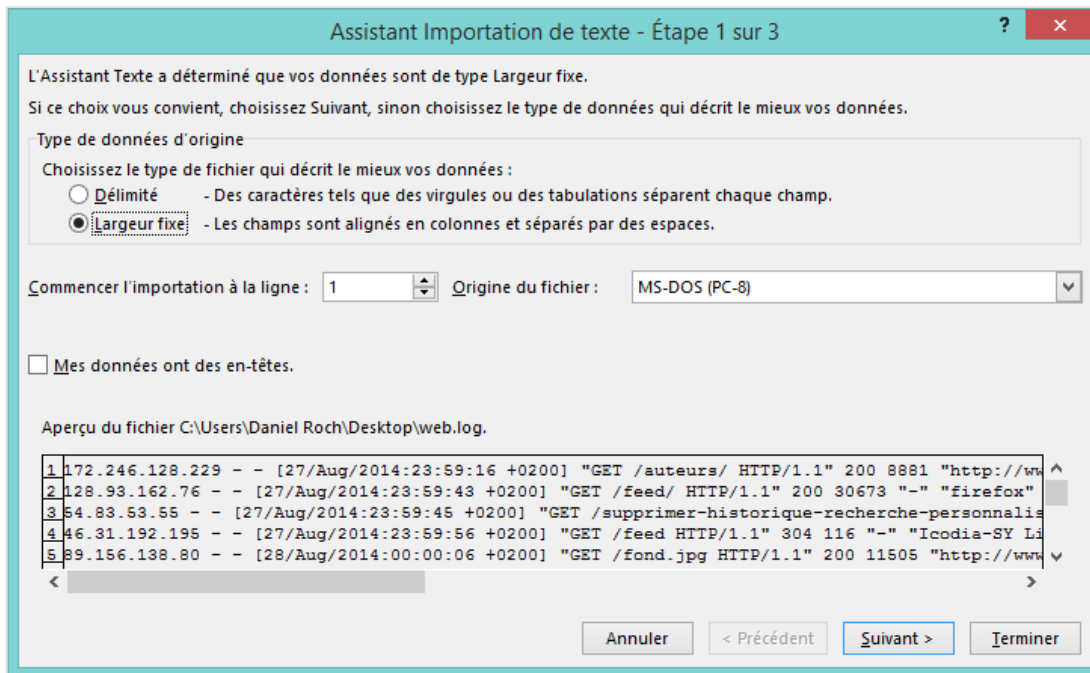
Attention cependant, ces scans ne pourront jamais faire ressortir les erreurs 404 provenant de liens erronés présents sur d'autres sites Internet. Ils ne traiteront que les problèmes de maillage interne.

L'analyse des logs serveurs

Il s'agit sans doute là de la source la plus pertinente d'informations, car ces données vont lister l'intégralité des erreurs 404 générées sur votre site.

L'idée est de télécharger les logs de votre serveur (demandez à votre hébergeur comment faire si vous n'y avez pas accès de façon simple). Une fois ce fichier de log en votre

possession, ouvrez un tableur Excel puis choisissez le menu d'importation des données. Dans Excel, il se trouvera dans « Données > Texte ». Sélectionnez alors votre fichier de log. Vous aurez ensuite une fenêtre qui ressemblera à celle-ci :



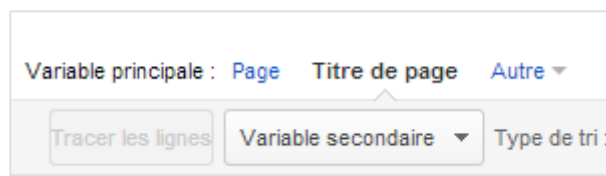
Choisissez l'option « délimitée » puis, dans le menu suivant, sélectionnez votre paramètre de délimitation de chaque donnée. Avec les logs de l'hébergeur Infomaniak par exemple, c'est le caractère ".

Libre à vous ensuite de trier la liste de vos logs pour ne garder que les erreurs 404, que ce soit celles générées par des robots (donc en partie par des moteurs de recherche) ou par des internautes.

Les outils de suivi

Vous avez peut-être installé un script de suivi pour vos visiteurs, votre positionnement ou encore pour les robots. Ces outils peuvent vous fournir également des informations sur vos pages d'erreurs.

Google Analytics par exemple peut vous fournir les erreurs 404 générées par un utilisateur. Pour cela, rendez-vous dans le menu « Comportement > Contenu du site > Toutes les pages », puis cliquez au-dessus du tableau sur la variable principale « Titre de page ».



Toujours en haut du tableau, sélectionnez la variable secondaire « Comportement > Page », puis utilisez le moteur de recherche en haut à droite du tableau et tapez « 404 », « erreur » ou encore « non trouvée » pour voir s'afficher vos pages d'erreurs.

Titre de page	Page	Pages vues	Consultations uniques	Temps moyen passé sur la page	Entrées	Taux de rebond
		324 % du total: 0,07 % (482 173)	285 % du total: 0,08 % (359 784)	00:01:02 Moyenne du site: 00:01:21 (-23,49 %)	178 % du total: 0,09 % (196 665)	67,98 % Moyenne du site: 62,52 % (8,74 %)
1. Page non trouvée - 404		36 (11,11 %)	20 (7,02 %)	00:00:44	13 (7,30 %)	23,08
2. Page non trouvée - 404		18 (5,56 %)	18 (6,32 %)	00:01:27	13 (7,30 %)	92,31
3. Page non trouvée - 404		15 (4,63 %)	13 (4,56 %)	00:00:12	12 (6,74 %)	83,33
4. Page non trouvée - 404		10 (3,09 %)	9 (3,16 %)	00:00:27	6 (3,37 %)	50,00
5. Page non trouvée - 404		10 (3,09 %)	9 (3,16 %)	00:00:19	1 (0,56 %)	0,00
6. Page non trouvée - 404		9 (2,78 %)	8 (2,81 %)	00:02:55	7 (3,93 %)	71,43

D'autres outils peuvent faire aussi cela de façon automatisée. Nous pensons notamment à :

- Les autres logiciels de webanalytics ;
- Watussi Box (pour analyser les logs) : <http://box.watussi.fr/> ;
- Pingdom Tools : <https://www.pingdom.com/> ;
- ...

Pensez donc à les consulter pour vérifier la présence ou non de pages 404.

Les outils internes

Dans certains cas de figure, l'outil que vous utilisez pour créer ou gérer votre site Web peut parfois vous aider. Si vous utilisez un CMS, il existe des plugins qui vont vous alerter en temps réel des erreurs 404 pour pouvoir ensuite les corriger. Par exemple sur WordPress, on pourra utiliser :

- Broken Link Checker <https://wordpress.org/plugins/broken-link-checker/> ;
- 404 Notifier <https://wordpress.org/plugins/404-notifier/> ;
- ...

Créer son propre système d'alerte

En fonction de votre serveur et de la façon dont est codé votre site, il vous est aussi possible de développer une fonction pour vous avertir en temps réel des erreurs 404 de votre site (en PHP, ASP...).

Dans ce cas, pensez à bien récupérer les informations suivantes pour que vous puissiez correctement les traiter par la suite :

- L'URL de la page d'erreur ;
- La date de l'erreur ;
- Le User-Agent (est-ce un internaute ou un robot ?) ;
- Le referrer (depuis quel contenu l'erreur a été générée).

Comment traiter les erreurs 404 ?

Corriger d'abord vos liens internes et externes

Quel que soit le type d'erreur rencontrée, la première étape est de faire en sorte que plus aucun lien ne pointe vers elle. Cette remarque est valable pour le SEO, mais également et avant tout pour le visiteur. Si vous voulez qu'il revienne, il faut à tout prix éviter l'orienter vers de "vrais" contenus.

Faites donc en sorte que plus aucun lien ne dirige les internautes vers ces pages, que ce soit des liens en interne ou en externe.

La seconde étape est alors de corriger la page d'erreur en elle-même, et il existe alors deux possibilités : la redirection 301 ou laisser un code HTTP 410 ou 404.

Une redirection 301

En règle générale, il est conseillé de rediriger les pages 404 vers vos autres contenus à l'aide d'une redirection 301 (permanente). Cela permet deux choses : conserver la popularité de la page initiale pour la transmettre à une autre, et éviter de donner une mauvaise expérience utilisateur à l'internaute en ne lui affichant pas une page d'erreur.

L'idéal est de toujours rediriger le visiteur vers le contenu le plus proche ou le plus pertinent par rapport à la page qu'il pouvait chercher initialement. Par exemple, si vous supprimez un article, redirigez le visiteur vers un autre article similaire ou vers sa catégorie.

En dernier recours, vous pourrez la rediriger vers l'accueil, mais sachez que cela reste la solution la moins pertinente de toutes : cette remarque est vraie pour le visiteur, mais aussi pour le moteur de recherche.

Pour faire une redirection 301, il existe plusieurs solutions :

- L'ajout d'une règle dans votre fichier HtAccess pour rediriger une URL à la fois ou un groupe d'URL ;
- L'utilisation d'un plugin ou outil externe, par exemple le plugin « Redirection » pour le CMS WordPress (<https://wordpress.org/plugins/redirection/>).

Un code 410 ou 404

Dans certains cas, il est conseillé de ne rien faire ou de renvoyer un code 410 « Contenu définitivement supprimé ». Google traite en effet de la même façon ces deux codes : <http://googlewebmastercentral.blogspot.fr/2011/05/do-404s-hurt-my-site.html>

Il existe quelques cas de figure où vous pouvez laisser vos pages 404 sans redirection. Par exemple :

- Plus aucun lien ne pointe vers cette page d'erreur (elle n'a donc aucune popularité et aucun visiteur ne reviendra dessus) ;
- Vous n'avez aucun contenu pertinent vers lequel rediriger l'internaute, et la page avait très peu de popularité.

En règle générale, on préférera la redirection 301, mais gardez toujours en tête que ce n'est pas toujours utile.

Le problème des softs 404

Les erreurs « soft 404 » sont assez récentes dans le centre pour webmasters de Google, et permettent de mettre en avant un type de contenu bien précis : les pages qui renvoient un code HTTP correct 200 et qui sont considérées par Google comme vides ou en erreur. En d'autres termes, ce sont toutes les pages mal conçues qui devraient renvoyer le code HTTP 404 mais qui ne le font pas.

Ces erreurs doivent donc être considérées au même titre qu'une erreur 404 classique, et vous devrez là encore corriger les liens qui dirigent vers elles puis choisir entre la redirection 301 ou le fait de renvoyer le code 404 ou 410 (source <https://support.google.com/webmasters/answer/181708?hl=fr>).

Au passage, quand vous rencontrez des erreurs de ce type, posez-vous toujours la question suivante : pourquoi mon site ne renvoie pas le bon code HTTP et comment faire pour corriger ce bug ?

Attention en ce qui concerne les erreurs en soft 404 indiquées par le centre pour webmasters de Google, la liste ne semble pas être exhaustive selon certains référenceurs (source <http://www.journaldunet.com/solutions/seo-referencement/seo-les-erreurs-dans-google-webmaster-tools/les-erreurs-soft-404.shtml>).

Conclusion

Les erreurs 404 posent parfois problème en référencement naturel en ralentissant votre indexation ou en occultant une partie de votre popularité. Pire encore, ces erreurs auront clairement un impact très négatif sur vos visiteurs.

Il est donc crucial de toujours analyser vos erreurs 404, que ce soit manuellement ou automatiquement, pour ensuite pouvoir les corriger de la meilleure façon possible.

Vous vous assurez ainsi de ne pas gêner votre référencement, mais surtout de donner une excellente expérience utilisateur à vos internautes. Et de gagner quelques précieux backlinks en passant !

Daniel Roch, *Consultant WordPress, Référencement et Webmarketing chez SeoMix (<http://www.seomix.fr/>)*