

Erreur 404 et SEO : ce qu'il faut savoir...



Par Aymeric Bouillat

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Tout site web doit gérer à un moment donné de son existence, et de façon plus ou moins forte, la problématique des erreurs 404 (page non disponible à l'URL demandée). En SEO, un travail doit également être effectué pour éviter un crawl inutile par les robots des moteurs, mais également pour récupérer des backlinks ainsi perdus. Bien qu'un trop grand nombre d'erreurs 404 ne génère aucune pénalité de la part de Google, malgré une croyance largement répandue, il est nécessaire de traiter avec efficacité ce problème, ne serait-ce que pour améliorer l'expérience utilisateur, gagner en popularité et améliorer son indexation.

Les erreurs 404 sont souvent problématiques pour le SEO : leur origine peut être diverse, et en fonction des pages impactées, les effets pour le trafic d'un site et son positionnement dans les pages de résultats des moteurs de recherche sont parfois non négligeables. La problématique des erreurs 404 touche aujourd'hui la majeure partie des sites web. Bien qu'il soit indispensable de les surveiller dans toute stratégie SEO pour optimiser le crawl et les performances globales du site, elles ne sont pas toujours néfastes et font partie du cycle de vie de tout site web. Nous allons voir dans cet article comment mieux les appréhender et analyser le degré de gravité de ces dernières pour le SEO.

Pourquoi il faut limiter leur nombre

Pour rappel, Google attribue un certain temps machine à chaque site en fonction de différents critères (autorité, profondeur du site, nombre total de pages, etc.). Plus un site aura d'URL qui répondent avec un code erreur 404 et qui répondait en 200 auparavant, plus Google passera de temps sur des pages qu'il risque de déclasser dans les pages de résultats et moins il se concentrera à crawler et analyser les pages pertinentes d'un site. **Mais, contrairement à un mythe courant dans le landerneau du SEO, il n'inflige pas directement une pénalité à un site, sous prétexte que celui-ci contient des erreurs 404,** comme expliqué dans cet article : <http://www.abondance.com/actualites/20150812-15406-404-et-penalites-backlinks-sur-des-soft-404-pipes-dans-les-url-quelques-infos-seo-sur-lalgorithme-google.html>

Les erreurs de crawl dans la Google Search Console (voir fig. 1) permettent de visualiser rapidement l'étendue des dégâts. Au-delà des problématiques que les erreurs 404 peuvent engendrer, il faut savoir qu'une URL qui renvoie un code 404 à plusieurs reprises disparaîtra des pages de résultats après quelques semaines, Google ne souhaitant pas proposer de contenus « introuvables » à ces utilisateurs, ce qui est somme toute assez logique.

Au-delà de l'aspect moteur, n'oublions pas qu'une URL renvoyant une erreur 404 peut aussi recevoir des liens en provenance d'autres sites : blogs, forums, réseaux sociaux, sites édités,... et donc avoir un effet négatif pour l'expérience utilisateur de visiteurs provenant de sites référents.



Figure 1. Variation du nombre d'erreurs 404 dans Google Search Console

Google peut d'ailleurs parfois remonter des erreurs 404 par rapport à des URL que ses crawlers auraient tenté de visiter, après avoir vu un élément s'apparentant à une URL dans les liens Javascript par exemple :

```
<a href="bonjour.pdf" onClick="_gaq.push(['_trackPageview','/télécharger-bonjour']);">PDF Bonjour</a>
```

Google perçoit « /télécharger-bonjour » comme étant un lien, qu'il va tenter de visiter, alors qu'il s'agit ici d'un marqueur pour Google Analytics. Cela provoquera l'affichage d'une page d'erreur 404 dans les « Erreurs d'exploration » de Google Search Console, une restriction dans le robots.txt dans ce type de cas corrigera le problème. Google appelle cela des « Erreurs 404 inattendues » dans sa documentation en ligne (<https://support.google.com/webmasters/answer/93641?hl=fr>), nous pourrions appeler ça « crawlers trop gourmands ! »

En cas de contenu supprimé de façon définitive, on peut également utiliser un code d'erreur 410 : « Gone ». Googlebot ira vérifier cette URL comme il le fait pour les 404, mais de façon de plus en plus espacée d'après nos constats. Cela ne signifie pas pour autant qu'il va s'arrêter de la crawler définitivement du jour au lendemain. Il ira parfois crawler à nouveau ces URL en 410 pour vérifier qu'elles n'existent vraiment plus et renvoient toujours le même code, dans le cas où elles recevraient des liens externes entre autre...

404, le netlinking facile

Utiliser des redirections 301 pour corriger des erreurs 404 est d'ailleurs un *quickwin* en terme de Netlinking. En effet, avant de chercher à développer la popularité de son site en obtenant de nouveaux liens, il est judicieux de capitaliser sur ses actifs hypertexte en allant chercher les liens cassés : il n'est pas rare de trouver des perles, parfois dans la même thématique :

- Sites faisant des liens vers des pages supprimées ;
- Sites faisant des liens erronés (encodage d'URL, copier/coller raté) ;
- Liens obsolètes (ex : migration de sites) non redirigés ;
- Erreurs de liens internes ;
- Boucles de liens / spider-traps.

Voici une méthodologie pour récupérer les liens cassés :

<http://blog.resoneo.com/2013/09/linkbuilding-capitalisez-sur-vos-actifs-hypertexte/>

Si les liens cassés trouvés font référence à une page qui n'existe plus, une redirection 301 vers une page similaire sera utile, si elle répond aux besoins de l'internaute qui cliquera sur le lien (« bien pour l'utilisateur, bien pour le SEO » comme nous le dit Google...). Attention cependant, tous les liens ne sont pas forcément bons à rediriger : ancrés sur-optimisés, sites hors thématiques, liens non pertinents, *mega-footer*, etc. La mise en place de redirection 301 de façon manuelle, au cas par cas, sera plus bénéfique, pour éviter Google Penguin ou une action manuelle.

Erreurs fréquentes

Les erreurs 404 sont souvent vues comme le point noir d'un site, alors que certaines d'entre elles ne sont pas forcément problématiques. Cette vision peut parfois conduire à des excès.

Redirections massives

Dans le cas d'une migration de site avec modification de la structure des URL par exemple, l'essentiel est de rediriger l'ensemble des URL vers les nouvelles URL correspondantes, pour ne pas déstabiliser l'internaute, mais aussi pour conserver la popularité et le positionnement de chacune des pages. Afin d'éviter un trop grand nombre d'erreurs 404 lors de la migration, si l'ensemble des URL n'ont pas été redirigées (faute de temps ou d'absence de correspondance), la facilité peut l'emporter sur la raison : rediriger toutes les URL qui n'ont pas de correspondance vers la page d'accueil... C'est une erreur !

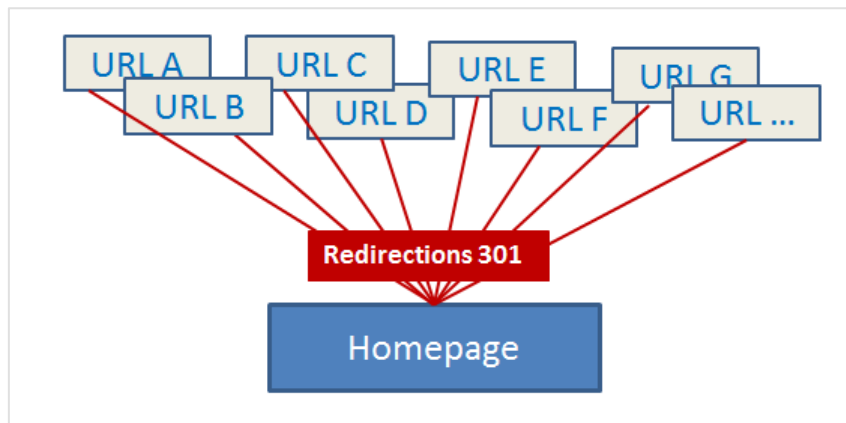


Fig. 2. Redirections massives pouvant être pénalisantes

Nous l'avons déjà constaté par le passé sur plusieurs sites, les redirections 301 massives de nombreuses URL vers une seule et même page peuvent s'apparenter à une manipulation du PageRank et de l'algorithme, une pénalité sur la page d'accueil peut donc se déclencher avec cette méthode peu recommandée. Cela engendrera d'autres problèmes puisque toute nouvelle erreur 404 (lien externe, nouvelle page supprimée) provoquera une redirection 301 vers la Homepage. Sans oublier l'expérience utilisateur qui risque de se trouver dégradée.

Au sujet des migrations SEO, il est d'ailleurs utile de nettoyer au mieux les erreurs 404 avant la migration, pour avoir une vue plus claire des 404 générés par les modifications d'URL après une mise en production. Vous trouverez sur cet article des informations pour découvrir et corriger les erreurs 404 :

<http://abonnes.abondance.com/archives/2014-09/RR-2014-09.htm>

Gestion des images

A l'inverse, certains éléments peuvent être oubliés et également générer des erreurs 404 : les images. Elles sont pourtant une source importante de trafic pour de nombreux sites, que ça soit via Google Images, ou les Onebox Images qui s'affichent directement dans les pages de résultats.

4,2 ★★★★★ avis sur conforama.fr
Canape Couleur en soldes Jusqu'à -50% de réductions !
[Soldes Conforama](#) - [Canapé d'Angle](#) - [Banquette Clic-Clac](#)

Images correspondant à canapé rouge Signaler des images inappropriées



Plus d'images pour canapé rouge

Canapé - Rouge - Achat en ligne - Vente Unique
www.vente-unique.com › [Canapé, Fauteuil](#) › [Vos canapés préférés](#) ▼
★★★★★ Note : 9,1/10 - 29 459 votes

Fig. 3. Onebox Image dans les pages de résultats Google

Le crawler Googlebot-Image ayant une fréquence de crawl moins importante que le crawler classique, la prise en compte des redirections 301 en cas de modification de l'URL d'une image (ex : nom de fichier) sera plus longue à être prise en compte. L'utilisation des logs serveurs sera un précieux atout de ce type de cas, les erreurs sur les fichiers images étant non trackées par les outils standards comme Google Analytics.

L'installation d'outils d'optimisation d'images sur un serveur Web peut également modifier les noms de fichiers médias, et provoquer des erreurs 404 lors du crawl de Googlebot-Image. LE module **ModPagespeed** pour Apache par exemple renommera [logo-seocampus-300x121.jpg](#) pour l'optimiser en [logo-seocampus-300x121.jpg.pagespeed.ic.B_VSrmdWjb.jpg](#). Si jamais vous désactivez ce module par la suite, l'URL contenant le nom de fichier modifié ne répondra plus. (Source : <http://www.yapasdequoi.com/apache/2962-modpagespeed-ou-lennemi-de-google-images.html>).

Soft 404 et mauvaises configurations

Il arrive parfois que des pages d'erreurs signalées comme étant de type « Page introuvable » (mention Erreur 404 dans le titre ou le contenu de la page par exemple) répondent en 200. Ces pages, appelées « Soft 404 », se retrouveront donc indexées et pollueront la liste de pages indexées par Google pour votre site. Il est nécessaire de vérifier dans ce cas si le CMS n'est pas mal paramétré ou n'interfère pas avec la configuration du serveur.

Nous constatons également parfois des redirections 302 qui précèdent la page d'erreur 404, ce qui n'est pas optimal pour la compréhension de Google et ce qui rendra le suivi plus complexe pour connaître les URL en 404.

Pour vérifier que votre site renvoie bien un status http 404, appelez une URL inexistante sur votre site : <http://votresite.com/blablablalatest> et vérifiez le code transmis dans l'en-tête HTTP de la page via l'inspecteur d'élément du navigateur, comme le montre la figure 4.

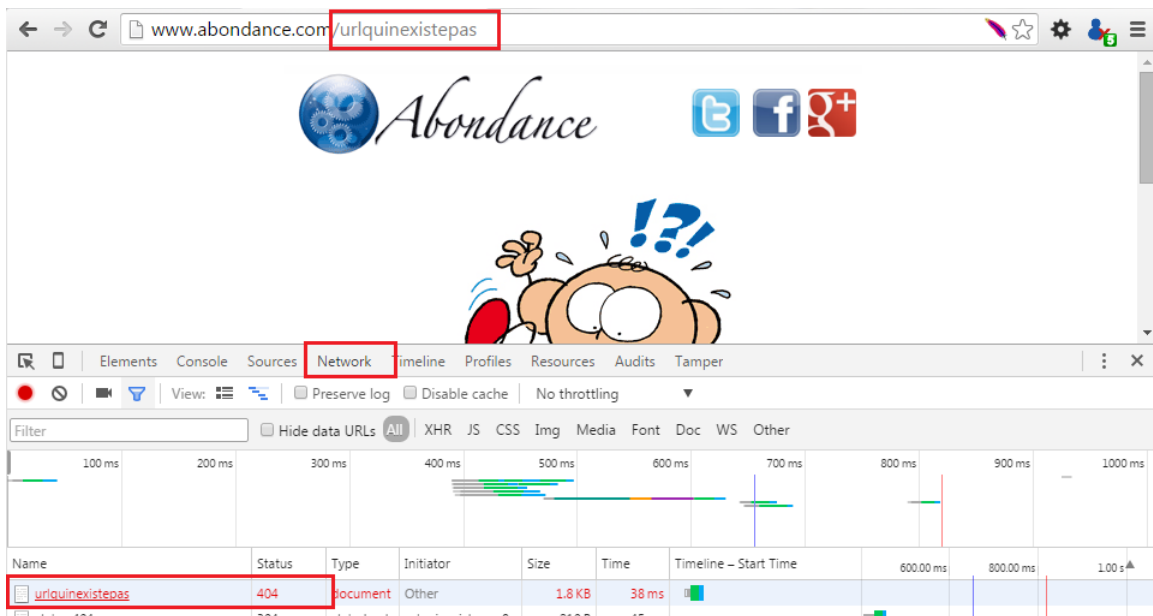


Fig. 4. Indication du code HTTP dans l'inspecteur d'élément du navigateur.

Tempérons leur impact

Comme mentionné précédemment, les erreurs 404 ne sont pas forcément toujours problématiques : elles peuvent concerner un nombre limité de pages, qui n'ont jamais été positionnées dans les moteurs de recherche, et qui ne pénaliseront donc pas le SEO pour autant. Au-delà des problématiques de crawl en cas de fort volume comme vu précédemment, le fait de corriger quelques erreurs 404 sur des pages peu pertinentes ne révolutionnera pas votre stratégie de référencement.

Il est nécessaire de trouver leur origine, pour parfaire sa stratégie SEO, mais dans certains cas, s'acharner à corriger des erreurs 404 sur des pages peu pertinentes sera une perte de temps par rapport à des actions qui auront plus d'impact comme l'optimisation du maillage interne, ou la correction de balises Title. Il vous faudra donc bien évaluer la gravité des 404 en fonction de leur courbe d'évolution, du type d'URL touchées, et de leur potentiel « trafic ».

404 et e-commerce, pas de solution miracle

Les erreurs 404 font partie du cycle de vie d'un site, elles sont parfois plus fréquentes pour les sites de e-commerce, il n'est pas rare de voir des produits supprimés répondre en 404. On pourrait faire l'analogie avec un grand magasin d'électroménager : vous vous rendez dans votre magasin préféré pour acheter un grille-pain, un modèle spécifique dont vous avez vu la référence sur plusieurs prospectus (i.e. des backlinks). Mais vous arrivez trop tard, ce modèle ne se fait plus, il n'est plus en stock (page d'erreur 404)... Allez-vous pour autant sortir de la boutique ? Quelle suite préféreriez-vous ?

- Etre redirigé par un vendeur vers la porte d'entrée, qui vous montrera tout ce qu'il propose dans ses rayons divers et variés (redirection 301 vers la homepage, assez inadaptée...);
- Qu'il vous propose un produit équivalent ? (redirection 301 vers un produit similaire);
- Qu'il vous présente la totalité du rayon grille-pain pour vous proposer d'autres produits de ce type (redirection 301 vers la catégorie mère);
- Qu'il vous dise clairement que le produit n'existe plus (erreur 404), et qu'il faut oublier cette référence (erreur 410).

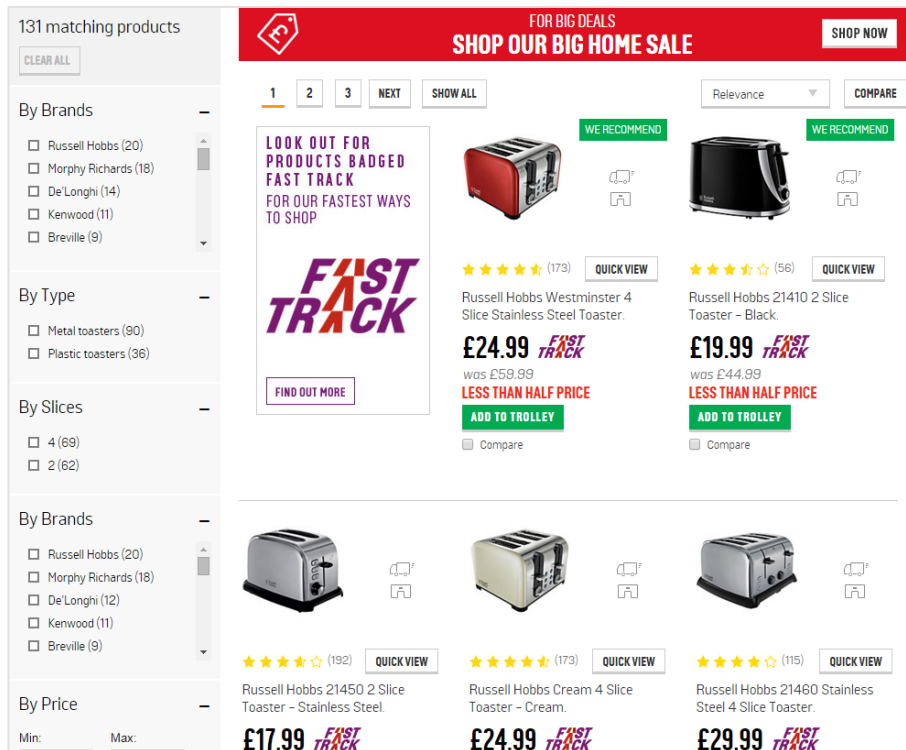


Fig. 5. Page catégorie d'un site de e-commerce

Aucune de ces solutions n'est applicable à 100% des sites e-commerce. En fonction de la structure des sites, du nombre de produits dans chaque catégorie, du cycle de renouvellement des gammes, et du volume des références en fin de vie, la stratégie peut être complètement différente...

Redirections 301 nombreuses

Bien que le fait d'avoir une redirection 301 vers un produit similaire ou la catégorie mère soit une solution adaptée dans de nombreux cas, elle peut parfois s'avérer contre-productive : on augmentera le nombre d'URL connues de Google, et donc le volume global d'URL à crawler (URL produits en augmentation, URL des anciens produits redirigés).

Si le volume de 301 devient trop élevé, on rentre dans une gestion complexe des redirections pouvant impacter la charge serveur (fichiers .htaccess volumineux), dont une grande partie d'entre elles n'auront que peu de poids au final : URL jamais positionnées, trop long-tail, transmettant peu de popularité, sur lesquelles Googlebot

passé une bonne partie de son temps (imaginer la mémoire de notre vendeur qui devrait connaître les produits similaires pour chacune de ses anciennes références depuis plusieurs années...)

Erreur 404 et perte de liens

On pourrait dans ce cas-là partir du principe que proposer une page d'erreur (404 ou 410) permettant à l'internaute de rebondir de lui-même vers une catégorie proche ou le moteur de recherche interne via une page d'erreur 404 personnalisée soit la solution. Mais dans le cas où plusieurs fiches produits reçoivent des liens externes, on perd une partie du netlinking : il faudra dans ce cas évaluer l'importance de la perte de popularité, pour décider de la stratégie à adopter, et gérer ce type d'URL au cas par cas (via un script connecté à l'API Ahrefs, ou MajesticSEO qui faciliterait le travail en semi-automatique).

Erreurs 404 inadaptées aux circonstances

Inutile par ailleurs de renvoyer un code d'erreur 404 sur une page proposant un produit temporairement indisponible sur un site de e-commerce: s'il n'y a plus de stock mais que le produit existe toujours, la page présentant le produit doit encore répondre en 200, avec un message spécifique permettant aux internautes de recevoir une alerte par mail dès que le produit sera à nouveau disponible. Une erreur 404 provoquerait une suppression de l'URL de l'index.

De trop nombreuses variations de status http 200 -> 404 ->200 -> 404... risqueraient également de réduire de façon plus importante la visibilité d'un site dans les pages de résultats, Google ne s'attardant probablement pas sur des pages non fiables à long terme...

Avoir une page d'erreur 404 personnalisée

Avoir une page d'erreur personnalisée reste un point important pour limiter le taux de rebond sur votre site avec une page de destination de ce type, sans « issue de secours ».

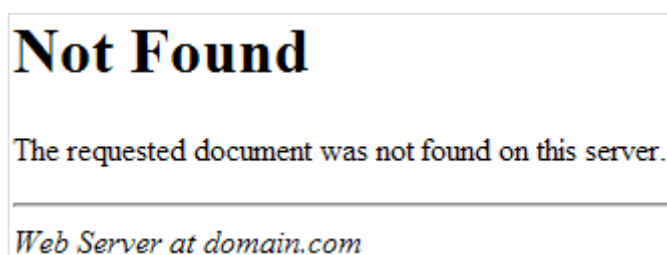


Fig. 6. Page d'erreur 404 non personnalisée

Une page personnalisée peut être composée de plusieurs éléments comme :

- Un message clair indiquant que la page n'existe plus ;

- Un lien vers la rubrique/catégorie de la page demandée (si la structure du site le permet) ;
- Un bloc présentant les derniers articles de la rubrique, ou des articles similaires de la catégorie ;
- Un accès direct au moteur de recherche interne pour trouver la réponse à sa requête initiale ;
- Un formulaire offrant la possibilité à vos visiteurs de vous signaler l'erreur qu'ils ont rencontré (URL en 404, URL contenant le lien générant une 404).

Elles peuvent également être originales comme certaines que vous trouverez sur ces sites qui regroupent les meilleurs du Web : <http://www.404notfound.fr/>, <http://fab404.com/>.

Pour personnaliser sa page d'erreur, il est possible de la modifier via le système de templates de la majeure partie des CMS du marché ([404.php](#) sur Wordpress, [page--error.tpl.php](#) sur Drupal, etc.). Dans le cas d'un site statique, c'est dans la configuration du serveur qu'il faudra indiquer son emplacement :

Pour Apache :

```
ErrorDocument 404 /404.html
```

Pour Nginx (section du fichier de configuration) :

```
error_page 404 /404.html;
```

Rendre l'accès à cette page impossible :

```
location /404.html {  
    internal;  
}
```

Mieux vaut prévenir que guérir

Pour suivre les URL renvoyant des erreurs 404 sur Google Analytics, il existe différentes méthodes qui sont décrites dans cet article :

<http://www.lunametrics.com/blog/2014/08/19/404-errors-google-analytics-google-tag-manager/>

La correction des erreurs 404 pouvant être une tâche qui devient rapidement chronophage, il est important de pouvoir analyser l'impact de ces erreurs (nombre, type d'URL touchées, origine des erreurs) pour prioriser leurs corrections afin d'améliorer votre référencement, et ce sans perdre du temps sur des redirections qui seraient inutiles.

Là où une seule erreur 404 sur une page très bien référencée peut s'avérer catastrophique, des centaines d'erreurs sur des pages n'ayant jamais existé (erreurs de liens) n'auront aucun impact. A vous de jouer !

Sources :

<https://httpd.apache.org/docs/2.4/fr/custom-error.html>

http://nginx.org/en/docs/http/nginx_http_core_module.html#error_page

<https://support.google.com/webmasters/answer/93641?hl=fr>



Aymeric Bouillat, consultant SEO, Resoneo (<http://twitter.com/aymerictwit> ou <http://www.yapasdequoi.com>)