

Le filtre « Safe Search » de Google rend-il une partie de votre site invisible ?



Par Philippe Yonnet

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Les contenus pornographiques sont partout sur le web, et les moteurs de recherche les indexent, au même titre que les autres contenus. Mais comment éviter que ces textes, ces images, ces vidéos, apparaissent dans les pages de résultats, notamment lors de recherches non "adultes", et surtout, quand ces recherches sont effectuées par des enfants ou des personnes susceptibles d'être choquées par ces contenus ? Et comment faire en sorte que vos contenus a priori innocents et loin de toute vision pornographique, ne tombent involontairement dans les affres des filtres de "Safe Search" mis en place par Google ?...

Très tôt, Google a déployé un filtre spécial, baptisé "Safe Search", pour faire face au défi des contenus "sensibles" et/ou pornographiques, et éviter les problèmes juridiques, mais aussi moraux, que ces contenus pouvaient créer dans une Amérique franchement puritaine.

Mais comment fonctionne ce filtre ? Quelle est son efficacité ? C'est que nous allons découvrir dans cet article. Nous verrons aussi que ce filtre peut rendre invisible pour la majorité des

internauts vos contenus, s'ils sont identifiés, à tort ou à raison, comme des contenus "adultes". Et que l'impact parfois très impressionnant de ce filtre peut même faire croire à l'existence d'une pénalité...

Qu'est-ce que le filtre safe search ?

Le filtre Safe Search est un processus du requêteur de Google (le composant qui génère la page de résultat en fonction

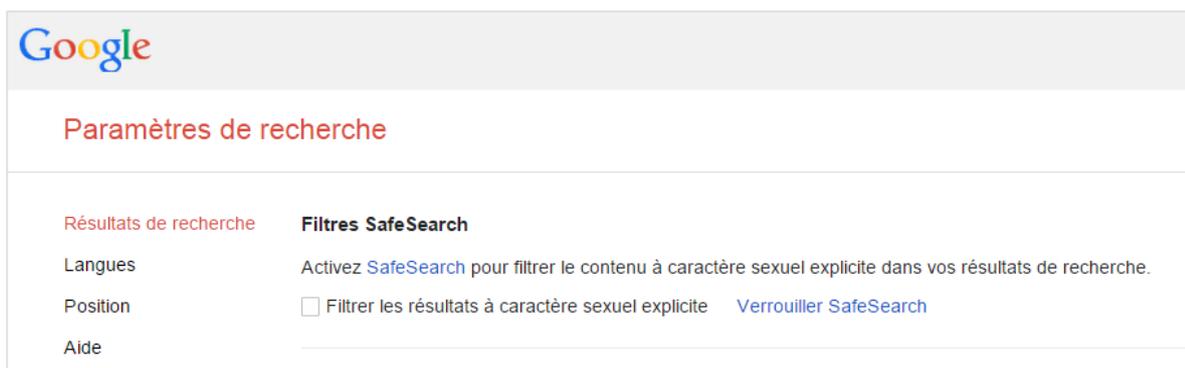


Fig.1. Prise en compte - ou non - du filtre "Safe Search" dans les préférences du moteur

de la requête de l'internaute) qui élimine des pages de résultats tout élément dont le contenu est classé "adulte" ou choquant : qu'il s'agisse de textes, d'images ou de vidéos. Ce filtre peut être activé ou désactivé via le menu paramètres de Google (voir figure 1 page précédente).

Avant 2012, il existait trois réglages pour ce filtre :

- *strict* : éliminait tout contenu adulte ;
- *modéré* : bloquait l'affichage des images pornographiques ou choquantes, mais pas les textes ;
- *off* : aucun filtrage.

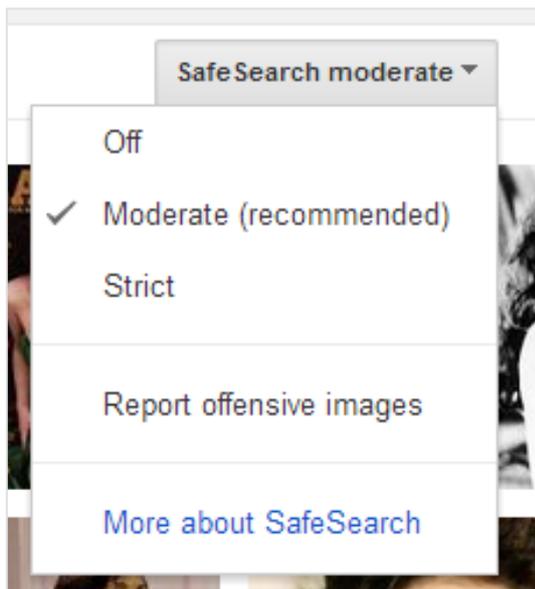


Fig. 2 : Les options proposées pour le Safesearch avant 2012

Le 12 décembre 2012, les 3 niveaux ont été remplacés par on/off sur les sites anglophones. Ce changement a été ensuite déployé sur de nombreuses autres versions locales de Google, y compris en France.

Par défaut, le filtre Safe Search était réglé à "modéré". En fait, c'est toujours le cas aujourd'hui, le réglage qui a

disparu est "off" : il n'est plus possible aujourd'hui de désactiver entièrement le filtre (sauf dans quelques rares pays dont la liste se raccourcit progressivement). Google n'offre donc plus la possibilité de "voir" certains résultats filtrés par Safe Search.

Depuis 2009, il est également possible de "verrouiller" Safe Search, afin d'activer le filtre pour tous les utilisateurs de votre profil ou de votre navigateur. Ce paramètre est activé depuis un compte google, et reste persistant par l'intermédiaire d'un cookie. Pour le désactiver, il faut se connecter au compte Google qui a verrouillé Safe Search, et changer le réglage. Cette fonctionnalité est utile en cas de partage d'un ordinateur familial avec des enfants par exemple.

Comment trouver du contenu pornographique avec Safe Search activé en mode modéré par défaut ?

En fait, Google a supprimé le mode "Safe Search off" parce que ses ingénieurs ont considéré comme plus sûr de détecter l'intention de l'utilisateur. Par exemple, dans Google Images, il suffit le plus souvent d'ajouter des termes comme "porn", "sex" ou "naked" à sa recherche, pour voir apparaître les contenus normalement filtrés comme par magie !

Comment fonctionne-t-il ?

Le filtre fonctionne différemment selon le type de contenu :



Fig.3. Un filtre safesearch verrouillé se reconnaît facilement par la présence de « boules bleues » sur la page de résultats. Ce signal visuel permet à des parents de vérifier d'un seul coup d'œil que le filtre strict est bien activé

- Pour le texte, il se base essentiellement sur la détection de mots clés "adultes" ou révélant un contenu choquant ou illicite. La présence de liens entrants ou sortants vers des sites ou des pages ou des ressources identifiées comme adultes ou choquantes est également prise en compte.
- Pour la vidéo, la catégorisation se fait en fonction des titres, des mots clés et des descriptions ajoutées sur les plateformes. L'analyse du contenu de la vidéo est également une piste de recherche
- Pour les images : la reconnaissance de forme est sollicitée et/ou une approche statistique de "détection de la peau". Une fois que l'on a reconnu dans une image les surfaces qui présentent les caractéristiques de coloration/texteure d'une peau, un traitement statistique probabiliste permet de détecter les



Fig. 4. L'image de Lena (une playmate) qui sert depuis de nombreuses années comme exemple dans la littérature scientifique en tant qu'image qui ne doit pas être reconnue comme pornographique... mais qui est classée comme telle par de nombreux algorithmes... Passer le « test de Lena » est donc un passage obligé pour valider la qualité d'un filtre « adulte ».

images susceptibles d'être pornographiques.

Comment connaître les mots clés filtrés ?

Un test simple suffit à identifier les mots clés "à risque" : tapez-les dans le champ de recherche en mode Safe Search activé, et si vous avez ce genre de message, c'est que ce terme est filtré :

*La fonctionnalité SafeSearch de Google étant activée, le terme "**blowjob**" a été ignoré lors de cette recherche.*

Les limites du filtre Safe Search

Comme Google l'avoue lui-même dans ses pages d'aide, le filtre Safe Search

n'est pas efficace à 100%. Il n'est pas rare donc de voir des images pornographiques ou choquantes apparaître sur des requêtes anodines. Mais les faux positifs sont également abondants.

La reconnaissance de mots clés crée des faux positifs, parce que les termes sont polysémiques (la polysémie est la caractéristique d'un mot ou d'une expression qui a plusieurs sens ou significations différentes : <http://fr.wikipedia.org/wiki/Polys%C3%A9mie>). On en veut pour preuve l'un des cas les plus extrêmes de faux positifs, qui est certainement celui du magazine canadien "The Beaver" (le Castor). Ce très sérieux magazine, fondé par la Compagnie de la Baie d'Hudson, est un magazine sur l'histoire



Fig.5. Un exemple de cas où le filtre a été pris en défaut...

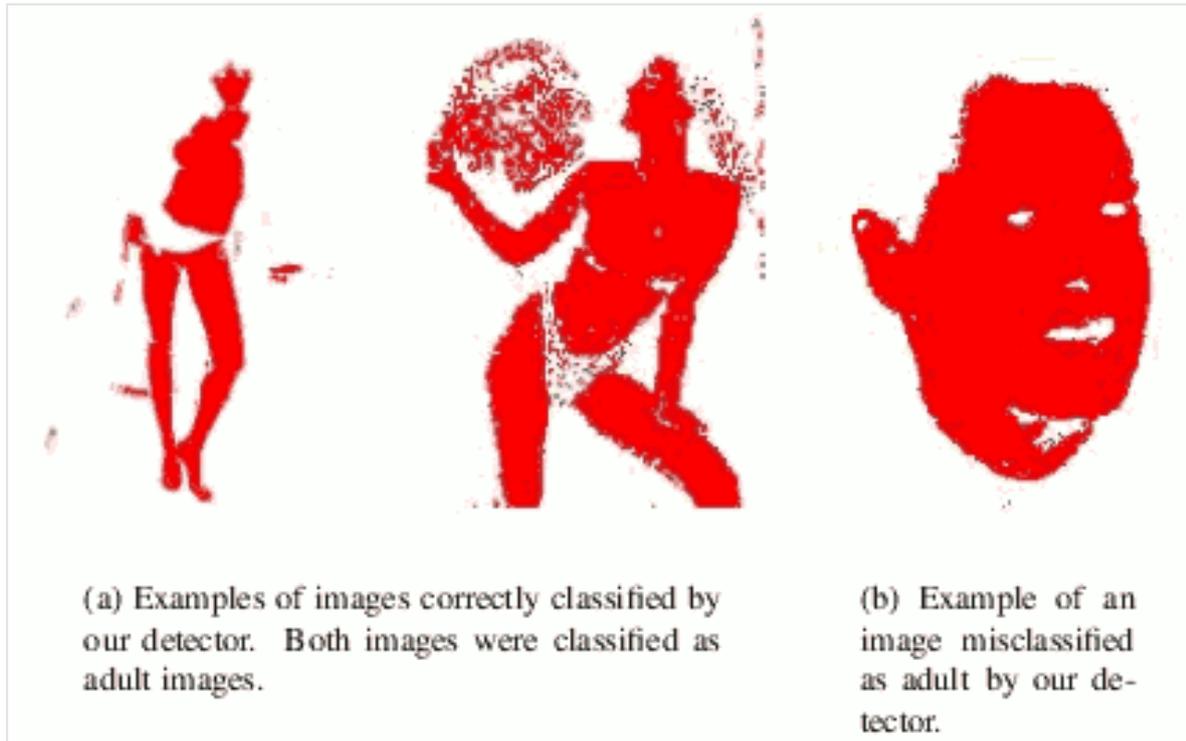


Fig.6. Autres exemples d'images plus ou moins bien identifiées...

du Canada. En 2010, le magazine a été contraint de changer son titre en "Canada's history" car tous les filtres anti spam empêchaient les enfants d'avoir accès à ses contenus, y compris et surtout via Google ! Beaver, au même titre que "chatte" en français, est un mot anglais à double sens !

Le filtrage des images pornographiques a lui aussi ses limites : la plupart des méthodes basées sur la détection de la peau ont tendance à filtrer toutes les photos de bébé comme pornographiques.

L'existence de ces faux positifs crée donc une menace potentielle pour la visibilité d'un site, car le filtre peut empêcher des pages anodines de se positionner...

Le filtrage par défaut et ses conséquences

Le réglage par défaut filtre les images et les vidéos apparaissant dans les pages de résultat. Dans la pratique, le filtre Safe Search peut avoir un impact très négatif sur le trafic issu de Google Images. Cet impact peut être très significatif. Et comme un pourcentage non négligeable active le filtre en mode strict, l'impact sur le trafic organique des pages web n'est pas neutre. Voici un exemple tiré d'un cas réel observé en 2009 sur un forum féminin :

- Pages du forum filtrées en mode strict : 800 000 sur 5 millions. Perte de trafic estimée : 15% ;
- Images filtrées par défaut : 680 000. Perte de trafic estimée depuis Google Images : 45%.

Comment détecter si vos contenus sont bel et bien considérés comme "adultes" ?

Comme il n'est plus possible de désactiver entièrement le filtre, détecter l'impact du filtre dans son réglage par défaut est devenu beaucoup plus difficile.

On peut néanmoins soupçonner fortement que le filtre Safe Search est à l'oeuvre dans les cas suivants :

- La ressource est crawlée (vérifiez dans les logs serveurs) ;
- La page ou la ressource ne reçoit pas de trafic moteur ou un trafic faible ;
- D'autres pages du même type (même template) se positionnent correctement ;
- Et il semble y avoir une forte corrélation entre contenu à risque et absence de trafic moteur !

Un moyen simple de confirmer ses soupçons est de tester le comportement de Bing avec le contenu suspecté d'être filtré. Bing dispose également d'un filtre Safesearch, qui fonctionne selon des critères proches de ceux de Google, mais le filtre de Bing a conservé son réglage à trois niveaux, et il est toujours possible de le désactiver. Si votre contenu apparaît dans les résultats de Bing avec le filtre Safe Search désactivé, et disparaît quand vous le réglez sur "modéré" ou "strict", vous aurez une première preuve que ce contenu est bien jugé "sensible".

Que faire pour éviter que vos contenus soient filtrés ?

Compte tenu de l'impact du filtre Safe Search, il est important d'éviter que son contenu soit abusivement considéré comme adulte.

Pour éviter que des pages tombent par erreur dans le filtre, il convient de prendre quelques précautions qui ont prouvé leur efficacité :

- Éviter l'emploi de certains mots clés "sensibles" dans les contenus, les noms de domaine, les titres de pages ;
- Éviter de présenter sur les pages des images potentiellement filtrées ;
- Éviter de faire des liens vers des contenus potentiellement filtrés.

Si un site comporte à la fois des contenus "adultes" ou "choquants" et des contenus "normaux", il convient de séparer les deux types de contenus pour éviter que TOUT le site disparaisse des résultats à cause du filtre Safe Search. Il faut donc :

- Placer les contenus "à risque" dans un répertoire séparé ;
- Limiter le crosslinking (maillage interne) entre pages "à risque" et pages "normales" au strict nécessaire.

Dans les cas extrêmes, on peut aussi :

- Placer une balise meta robots avec la valeur d'attribut "noindex" sur ce contenu ;
- Voire bloquer l'exploration des pages à risque par une directive du fichier robots.txt .

Mais dans ce cas, les pages à risque n'ont aucune chance de se positionner, donc c'est une méthode à n'utiliser que quand les autres ont échoué.

Pensez à vérifier l'impact du filtre Safe Search sur votre visibilité

L'impact du filtre Safe Search sur la visibilité d'un site est donc potentiellement important. Si vous présentez un type de contenu (images, vidéos, textes) qui est susceptible d'être filtré par SafeSearch, il faudra vérifier si ce filtre n'a pas un impact sur votre référencement. En particulier, si ces pages ne sont pas "actives" (si elles ne reçoivent pas de trafic en provenance des moteurs de recherche), cela peut être l'effet direct de SafeSearch, et non d'une pénalité. Pensez-y la prochaine fois que vous auditez un site avec ce type de contenu !

Bibliographie

Une étude de la Harvard Law School sur le filtre Safe Search datant de 2003 : http://cyber.law.harvard.edu/archived_content/people/edelman/google-safesearch/

Un papier scientifique d'un chercheur de Google sur le filtrage des images : *LARGE SCALE IMAGE-BASED ADULT-CONTENT FILTERING (2006)*

Henry A. Rowley, Google, Inc., Mountain View, CA, USA

Yushi Jing

College of Computing, Georgia Institute of Technology, Atlanta, GA, USA

Shumeet Baluja

Google, Inc., Mountain View, CA, USA

Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA

<http://static.googleusercontent.com/media/research.google.com/fr//pubs/archive/38.pdf>

et sur le filtrage par catégorisation des vidéos :

YouTubeCat: Learning to Categorize Wild Web Videos

Zheshen Wang, Ming Zhao, Yang Song, Sanjiv Kumar and Baoxin Li

<http://static.googleusercontent.com/media/research.google.com/fr//pubs/archive/36387.pdf>



Philippe YONNET, Directeur Général de l'agence *Search-Foresight*, groupe *My Media* (<http://www.search-foresight.com>).