

Comment les moteurs référencent-ils les pages sécurisées (https) ?

[Retour au sommaire de la lettre](#)

Nous vous proposons dans cette rubrique une série de "trucs et astuces" qui devraient vous simplifier la vie dans le cadre du référencement de votre site web. Ce mois-ci : le fait qu'une page web se trouve dans un espace sécurisé (adresse commençant par "https:") est-il un frein au référencement ? Réponse pour Google, Yahoo! Search et MSN Search.

De nombreux sites, sur le Web, proposent un espace sécurisé, soit pour leurs clients, soit pour saisir un numéro de carte bancaire, soit pour d'autres raisons. S'il est logique que des pages qui contiennent des informations personnelles (accessibles, par exemple, sur saisie d'un mot de passe) ne soient pas accessibles aux "spiders" des moteurs, on peut, en revanche, se poser la question de pages d'information comme celles qui présentent la solution "AdWords" de Google, présentes à l'adresse :

<https://adwords.google.fr/select/>

Ce sont des pages d'information, non confidentielles, qui ont tout à fait leur place dans l'index d'un moteur. Toute la partie "Description et FAQ" de l'offre commerciale se trouve, par exemple, dans l'espace sécurisé :

<https://adwords.google.fr/support/?hl=fr>

Il serait logique que ces pages, qui n'ont rien de personnel par rapport à l'internaute, se retrouvent dans les index des moteurs. Mais ceux-ci les indexent-ils ? Le fait que les adresses de ces pages soient en "https" est-il un frein ou un blocage à leur indexation ou leur positionnement ? C'est ce que nous allons voir, moteur par moteur.

Google

A priori, pas de soucis pour Google, il indexe parfaitement les pages de son site :

<http://www.google.fr/search?q=site:adwords.google.fr>

De même, le site web du Gouvernement français concernant la protection des mineurs (il contient de très nombreuses pages en zone sécurisée) est, lui aussi, référencé :

<http://www.google.fr/search?q=site:internet-mineurs.gouv.fr>

Enfin, troisième exemple, un site du Crédit Lyonnais développé sous cette forme est également indexé :

<http://www.google.fr/search?q=site%3Ainteractif.creditlyonnais.fr>

Il ne semble donc pas que le fait qu'une page ait une adresse en "https" soit un frein à son indexation. Mais qu'en est-il pour le positionnement ? Reprenons nos exemples... Pour le mot clé "AdWords", sur Google, c'est la page officielle qui est affichée en premier (sur 90 600 résultats) :

http://www.google.fr/search?hl=fr&c2coff=1&q=adwords&meta=lr=lang_fr

Idem pour la requête "protection des mineurs". C'est le site ministériel qui ressort en premier sur 806 000 réponses :

<http://www.google.fr/search?hl=fr&c2coff=1&q=protection+des+mineurs&spell=1>

Toujours pareil pour la requête "Crédit Lyonnais Interactif" : le site sécurisé est premier sur 12 200 réponses :

<http://www.google.fr/search?q=cr%C3%A9dit+lyonnais+interactif>

Donc, a priori, les pages issues de zones sécurisées semblent plutôt bien prises en compte par Google.

Yahoo! Search

Si Google semble indifférent au fait qu'une page soit ou non en zone sécurisée, qu'en est-il pour Yahoo! ?

De la même façon, sur la requête "site:adwords.google.fr", Yahoo! renvoie un résultat mais uniquement la page d'accueil :

<http://fr.search.yahoo.com/search?p=site:adwords.google.fr>

En revanche, le site du Gouvernement est bien présent avec de nombreuses pages :

<http://fr.search.yahoo.com/search?p=site:internet-mineurs.gouv.fr>

Tout comme le site du Crédit Lyonnais :

<http://fr.search.yahoo.com/search?p=site%3Ainteractif.creditlyonnais.fr>

De la même façon, le site officiel de Gogle est en tête de liste pour la requête "adwords" :

<http://fr.search.yahoo.com/search?p=adwords>

Tout comme le site du Gouvernement français pour "protection des mineurs" :

<http://fr.search.yahoo.com/search?p=protection+des+mineurs>

Et le site de la banque pour "crédit lyonnais interactif" :

<http://fr.search.yahoo.com/search?p=Cr%C3%A9dit+lyonnais+interactif>

Peu de différence, donc, à première vue, entre Google et Yahoo!. Les pages sont indexées et bien classées, quel que soit leur nature, sécurisée ou non. Qu'en est-il de MSN maintenant ?

MSN Search

De la même façon, la requête "site:adwords.google.fr" identifie le site de Google mais, cette fois, sous sa forme non sécurisée :

<http://search.msn.fr/results.aspx?q=site:adwords.google.fr>



Même chose pour le site gouvernemental : seules les pages en version non sécurisée sont identifiées :

<http://search.msn.fr/results.aspx?q=site%3ainternet-mineurs.gouv.fr>

Idem pour le Crédit Lyonnais :

<http://search.msn.fr/results.aspx?q=site%3Ainteractif.creditlyonnais.fr>

En revanche, plusieurs autres tests sur des sites web de notre connaissance, disponibles uniquement en version sécurisées, se sont révélés négatifs : aucun d'entre eux n'ont pu être identifiés dans l'index de MSN.

Il semble donc que ce moteur n'indexe pas - pour l'instant - les pages disposant d'une URL en "https"... Ce qui règle le problème de leur positionnement :-)

Conclusion

A l'heure actuelle, les pages présentes en zone sécurisée ne semblent présenter de problèmes ni pour Google, ni pour Yahoo!. En revanche, MSN semble les ignorer. Est-ce partie remise uniquement, au vu de la relative jeunesse de cette technologie ? Cela reste à voir.

Attention, cependant, de ne pas tirer de conclusions trop hâtives pour Google et Yahoo! : si, selon nos tests, ce type de page ne semble pas leur poser de problèmes, certains référenceurs nous ont fait part de difficultés, dans le passé, pour indexer et positionner les pages de certains sites présentant cette particularité. Prudence donc, un test préalable peut s'avérer indispensable selon le site pris en compte... A ce sujet, plusieurs points plus précis nous ont été fournis par ces sociétés de référencement. Nous vous les livrons "tels quels", car ils sont parfois complexes à vérifier techniquement :

- Il semblerait que les moteurs comme Google et Yahoo! aient pris comme stratégie d'indexer les pages web en "https", uniquement si le site propose également des pages en "http" par ailleurs. En d'autres termes, si un site est accessible uniquement et exclusivement en "https", sans aucune page disponible en version non sécurisée, cela peut poser des problèmes d'indexation. En tout état de cause, nous vous conseillons de créer des versions non sécurisées, si cela vous est possible, de vos pages d'information en "https" afin de faciliter l'indexation et le positionnement de votre site sur tous les moteurs et notamment MSN Search.

- Le fait qu'une page soit en "https" peut poser des problèmes au niveau du calcul du PageRank de Google. En effet, le "s" supplémentaire est parfois oublié dans les liens qui pointent vers le site. Donc, si la version non-sécurisée du site existe, c'est elle qui profite du lien, et non pas la version sécurisée. Du coup, une page en "https" aurait un risque de popularité moins grande qu'une page en "http"... Le fait de créer une version non sécurisée (http) évite également une erreur 404 sur un lien au sein duquel le "s" aurait été oublié...

- Les nouvelles versions de la barre d'outils de Google ne donneraient pas de PageRank aux pages en "https". Exemple : l'adresse **https://www.internet-mineurs.gouv.fr/** a un PageRank inconnu de Google, alors que **http://www.internet-mineurs.gouv.fr/** affiche un PR de 6. Cela veut-il dire que les pages sécurisées sont moins bien prises en compte par ce moteur ? Mystère...

- Certains hébergeurs peuvent bloquer les robots, par exemple à l'aide du fichier httpd (Apache), pour l'accès des pages sécurisées, et ce pour plusieurs raisons (sécurisation, charge serveur, etc.). Vérifiez, donc, auprès de votre hébergeur si ce n'est pas votre cas...

- Certains moteurs peuvent également vérifier le certificat du site : est-il expiré ? Qui l'a émis ? On peut imaginer que le moteur refuse une page correspondant à un certificat dont la date de validité a expiré. De la même façon, il se peut que le robot refuse d'indexer une page sécurisée dont le certificat n'aurait pas été émis par une autorité de confiance reconnue. Exemple, si c'est Verisign qui l'a émis, l'indexation est effectuée, mais si l'émetteur est inconnu, cela peut poser problème.

Notons, pour terminer, que les pages sécurisées disponibles et visibles uniquement sur saisie d'un mot de passe ne seront bien sûr pas indexées par les moteurs. Les spiders n'ont pas, pour leur part, connaissance de ces mots de passe. Ouf... ;-)