

Referer et vie privée

Big brother is watching you ?

Ce - bref je l'espère - article va nous permettre de voir ce qu'est ce que l'on appelle, sur Internet, le "Referer". Son but premier est de répondre à une question qui m'a été posée à l'occasion d'un cours en rapport avec les technologies et stratégies relatives aux sites de commerce électronique ; et je me suis dit qu'en l'étroffant un peu, cette réponse pouvait intéresser plus d'une personne, ce qui explique que je la diffuse sur ce site.

I:\ Qu'est-ce que le referer ?

Tout d'abord, je pense que quelques mots sur ce qu'est le "Referer" peuvent être utiles... Nous commencerons par une définition théorique, puis nous montrerons par un exemple technique ce dont il s'agit, et nous finirons, pour ceux d'entre vous qui n'ont pas de notion avancée en standards d'Internet, par une explication plus simple et plus concise.

A: Définition

Pour commencer, rien de mieux qu'un passage de la [RFC n°2616](#), qui définit le protocole HTTP/1.1, utilisé notamment pour le transfert de la plupart des pages web sur Internet ; ce passage correspond à un extrait de la section 14.36, qui définit ce qu'est le Referer, et qui fixe la norme à laquelle il doit se plier :

14.36 Referer

```
The Referer[sic] request-header field allows the client to specify, for the server's benefit, the address (URI) of the resource from which the Request-URI was obtained (the "referrer", although the header field is misspelled.) The Referer request-header allows a server to generate lists of back-links to resources for interest, logging, optimized caching, etc. It also allows obsolete or mistyped links to be traced for maintenance. The Referer field MUST NOT be sent if the Request-URI was obtained from a source that does not have its own URI, such as input from the user keyboard.
```

La traduction en français, par mes soins, donne ceci :

Le champ Referer de l'entête d'une requête permet au client de spécifier, au bénéfice du serveur, l'adresse (URI) de la ressource depuis laquelle l'adresse correspondant à la requête (Le "referrer", bien que le champ d'entête ait été incorrectement orthographié.). Le champ Referer de l'entête de la requête permet à un serveur de générer une liste de liens arrière vers des "sites intéressants" ou de les logger, d'optimiser la mise en cache, ... Il permet aussi aux liens obsolètes ou mal-saisis d'être tracés en vue de maintenance. Le champ Referer ne DOIT PAS être envoyé si l'adresse de la requête a été obtenue d'une source qui n'a pas sa propre adresse, telle une saisie effectuée au clavier par l'utilisateur.

B: En pratique

Prenons un exemple simple : je recherche de la documentation en rapport avec la programmation en Perl ; ne connaissant pas d'adresse de site spécialisé dans ce domaine, je décide de passer par un moteur de recherche ; [google.fr](http://www.google.fr) dans notre exemple.

Je choisis de rechercher "programmation perl" (sans les guillemets), et de me restreindre aux "Pages francophones", et je lance la recherche. (Pour une recherche concernant de l'informatique, il aurait peut-être mieux, dans d'autres cas, faire une recherche à l'international... mais pour le Perl, je sais que cette requête me donnera, parmi les premiers résultats, un lien vers un Tutorial que j'ai déjà utilisé ^^).

Google me propose, en troisième réponse, [Sylvain Lhullier : Publications : Introduction à Perl](#). Ce site me paraît intéressant (il l'est, mais ce n'est pas le sujet ici) ; je clique donc sur ce lien.

Lorsque je demande la page, mon navigateur envoie la requête HTTP suivante :

```
GET /publications/perl.html HTTP/1.1
Host: sylvain.lhullier.org
User-Agent: Mozilla/5.0 (X11; U; Linux i686; en-US; rv:1.7.12) Gecko/20050922
Firefox/1.0.7 (Debian package 1.0.7-1)
Accept:
text/xml,application/xml,application/xhtml+xml,text/html;q=0.9,text/plain;q=0.8,image/png
,*/*;q=0.5
Accept-Language: fr-fr,en-us;q=0.7,en;q=0.3
Accept-Encoding: gzip,deflate
Accept-Charset: ISO-8859-15,utf-8;q=0.7,*;q=0.7
Keep-Alive: 300
Connection: keep-alive
Referer:
http://www.google.fr/search?hl=fr&q=programmation+perl&btnG=Recherche+Google&meta=1r%3Dla
ng_fr
```

La ligne "GET" correspond à la demande de la page dont l'adresse est précisée, relativement à l'entrée "Host".

La ligne "User-Agent" correspond au navigateur que l'utilisateur utilise ; par exemple, ici, on voit que j'utilise Firefox en sa version 1.0.7, que je suis sous Linux, et que la distribution que j'utilise est Debian. Les différents "Accept" correspondent à différents paramètres du navigateur, tels les langues que l'utilisateur connaît, les charsets qu'il admet, ... Toutes ces informations sont destinées principalement à permettre au site de m'envoyer en réponse la page qui me convient le mieux ; par exemple, le site peut se rendre compte que je préfère le français à l'anglais, et, s'il en a la possibilité, m'envoyer la version française de la page que je demande.

Enfin, la ligne "Referer" correspond au sujet de l'article que vous êtes en train de lire : à l'aide de cette information le site sera à même de déterminer que c'est google qui m'a permis d'arriver à lui, via la recherche des termes "programmation perl".

C: Plus simplement...

Pour faire bref, le "Referer" est une information qui peut être transmise par votre navigateur Web lorsque vous demandez une page d'un site, qui indique par quelle page vous êtes arrivé sur celle que vous demandez.

II:\ Avantages et Dangers du Referer

Certains vous dirons que le Referer est indispensable ; d'autres, qu'il est utile ; et les derniers vous dirons que le Referer est une abomination...

Au cours de cette partie, nous verrons en quoi le Referer peut être utile pour le webmaster ; nous continuerons en étudiant quelques uns des services qu'il peut rendre au visiteur du site, et nous finirons par voir quelques uns des "dangers" qu'il peut représenter.

A: Avantages pour le webmaster

Je pense qu'aucun webmaster ne pourra vous dire que le Referer est une mauvaise chose, étant donné les services qu'il est à même de rendre. Cette partie va nous permettre d'en voir quelques uns.

Tout d'abord, le Referer est un élément qui est très souvent pris en compte dans les statistiques d'un site : il permet de savoir quels sont les sites qui rapportent des visiteurs, et donc, quels sont les partenariats et échanges de liens qui marchent bien. A partir de ces informations, le webmaster peut choisir de privilégier des relations avec certains de ses partenaires, en écartant éventuellement ceux qui ne lui apportent que peu de visiteurs.

Ensuite, connaître les Referer ayant permis d'accéder à son site peut permettre au webmaster de savoir lorsque quelqu'un parle de son site (poste un lien vers celui-ci) ailleurs - sur un forum, par exemple. Cela peut lui permettre de réaliser des améliorations suite à des critiques ou remarques qui ont été faites en public au sujet de son site, mais qui ne lui ont pas été directement adressées.

Enfin, dans certaines situations, le Referer permet de vérifier que l'utilisateur qui accède à une page vient bien de la page qu'il faut.

Ceci s'applique tout particulièrement dans le cas d'une page de gestion des données d'un formulaire : une des vérifications que l'on peut faire est de s'assurer que la page d'où vient le visiteur est celle qui correspond au formulaire en lui-même.

Cette technique est par exemple utilisée pour certaines options d'administration du moteur de forum de PunBB ; Cf le passage de la [FAQ de punBB en rapport avec le Referer](#).

B: Avantages pour le visiteur

Tout d'abord, notons que le but premier d'un webmaster est souvent, surtout sur un site non-commercial, de faciliter la navigation sur son site ; autrement dit de vous simplifier la vie, à vous visiteurs. Et comme nous venons de le montrer, pour ce genre de choses, le Referer peut être utile.

Ensuite, notons que certains sites ne fonctionnent pas, ou mal, sans le Referer ; on retombe sur le dernier point dont j'ai parlé dans la sous-partie précédente, qui traitait du fait que le Referer est parfois utilisé comme une sécurité.

Enfin, et c'est peut-être le plus important, ou, du moins, le plus visible par l'internaute, certains sites utilisent le Referer pour fournir des fonctionnalités supplémentaires à leurs visiteurs.

Par exemple, il existe des sites (et celui sur lequel vous vous trouvez en ce moment en fait parti !) qui détectent lorsque l'utilisateur est arrivé depuis un des grands moteurs de recherche, et qui mettent en valeur les mots qui ont été recherchés sur ce moteur ; de la sorte, le visiteur peut plus facilement, en théorie, trouver sur la page du site le ou les paragraphes en rapport avec le sujet de sa recherche.

C: Dangers

Comme toute technologie, le Referer peut être exploité de manière qui soit possiblement préjudiciable à l'utilisateur... ou du moins, il existe des utilisateurs qui ont peur que ce soit le cas.

Tout d'abord, comme déjà vu, il permet de suivre la navigation d'un internaute sur un site... cela dit, si vous visitez le site, il est possible pour le webmaster de vous suivre, même sans referer, puisque vous demandez de toute façon au serveur les pages que vous souhaitez visiter...

Ensuite, il peut y avoir des sites, ou des documents, que vous ne voulez pas voir transmis en Referer ; notamment, il peut paraître compréhensible que vous ne souhaitiez pas voir apparaître en Referer l'adresse d'un document présent sur votre disque dur (tel, par exemple, une page d'un site que vous auriez enregistré pour la consulter hors-ligne, et sur laquelle vous auriez cliqué sur un lien), ou encore une adresse correspondant à un formulaire de paiement en ligne - ce afin de ne pas exposer votre arborescence de répertoire dans le premier cas, ou ne pas indiquer au webmaster d'un autre site que vous venez d'effectuer un achat en ligne (Celui-ci pourrait alors se dire que vous êtes un bon client potentiel, s'il est à la tête d'un site proposant des produits à la vente). Cela dit, nous verrons au cours de notre sous-partie suivante que ceci n'est pas réellement un problème.

Vous tapez - par exemple - 'sexe' ou 'xxx' dans un moteur de recherche, ce qui vous permet d'arriver sur un site... et vous auriez voulu que celui-ci ne sache pas quelle est la recherche que vous aviez fait... après tout, pourquoi pas ? C'est un cas de figure qui se produit régulièrement.

Cela dit, si vous visitez un site suite à une recherche sur un moteur de recherche, je ne vois pas vraiment l'utilité de cacher la recherche que vous avez fait, puisque de toute façon vous êtes sur le site : de toute façon, le webmaster sait que vous l'avez visité... Alors, quel préjudice le fait qu'il sache comment vous y êtes arrivé pourrait-il vous apporter ?

D'un autre côté, si vous arrivez sur un site traitant d'un sujet sans aucun rapport avec celui de votre recherche, permettre au webmaster de connaître la recherche que vous avez fait peut lui permettre de modifier son site, afin d'essayer de ne plus apparaître dans les moteurs de recherche pour celle-ci... et donc, que les internautes aient moins d'informations parasites lors d'une recherche...

En somme, ici, laisser le Referer activé ne vous fait aucun mal... et peut aider le webmaster... ce qui au final est une bonne chose pour ses visiteurs !

III:\ (Dé)bloquer le Referer ?

A: Sécurité... quand tu nous tiens

Nous avons dit un peu plus haut que le Referer pouvait constituer un risque pour la sécurité dans certains cas, notamment en exposant des adresses de fichiers présents sur votre disque dur (file://), ou des adresses de pages sécurisées (https://).

Cela dit, dans ce genre de situation, le Referer ne constitue pas réellement un risque...

En effet, comme nous l'indique [ce bulletin Microsoft](#), Internet Explorer, le navigateur actuellement le plus répandu, ne transmet pas le Referer dans les situations considérées comme "non-secure"... Et comme j'ai pu le constater, [Firefox](#), le navigateur libre qui monte en puissance (et que je vous conseille !), agit de la même manière.

En somme, ici, c'est votre navigateur qui protège certaines informations que vous pourriez considérer comme sensibles.

B: Firewall et Referer

Selon votre politique de sécurité, ou celle de votre entreprise si vous naviguez sur Internet depuis votre lieu de travail, un logiciel de type Firewall ("pare-feu" en français) a pu être installé sur votre machine, en vue de fermer quelques portes au nez d'éventuels petits malins mal-intentionnés.

Or, certains Firewall se chargent, dans leur configuration par défaut, de bloquer le Referer. C'est par exemple le cas pour Norton Internet Security ou Norton Personal Firewall, [versions 2002 et précédentes](#) et [versions 2003 et supérieures](#), ainsi que pour Kerio, un Firewall gratuit pour les particuliers (voir cette [page du manuel de Kerio](#)). Il se peut tout à fait que ce soit aussi le cas pour d'autres logiciels de ce type.

Certes, nous avons vu que, dans quelques rares cas, le Referer peut être considéré comme un danger pour l'utilisateur... Mais nous avons aussi démontré qu'il était utile, voire même nécessaire, dans un certain nombre de situations. Il se peut donc que ces logiciels de "sécurité", dans leur configuration par défaut, puissent nuire à votre navigation sur Internet !

Autrement dit, en cas de problème... il peut être une bonne idée que de faire un tour au sein de la configuration de vos logiciels...

C: Jouons avec le Referer !

Avant que nous n'arrivions à la fin de cet article, je voudrais encore une fois attirer votre attention, surtout si vous êtes webmaster, sur le fait que **le Referer est une information qui n'est pas fiable**.

En effet, nous avons déjà vu que, dans certains cas, le navigateur choisissait de ne pas le transmettre, et qu'il existait des logiciels de sécurité qui le bloquaient...

Mais il est aussi possible de le fausser ! Oui ! Même si cela ne sert pas à grand chose en pratique, il est possible pour l'utilisateur de demander une page en envoyant le Referer qu'elle souhaite.

Notamment, il existe des plug-in qui s'intègrent au sein du navigateur Firefox qui font cela.

Par exemple, on peut citer [le plug-in RefSpoof](#), qui vous permet de demander une page en spécifiant absolument ce que vous voulez comme Referer.

Pour le développeur-web, cela peut permettre de tester certaines fonctionnalités de son site, en faire croire à celui qu'il vient, par exemple, d'un moteur de recherche. Pour le visiteur lambda, ça ne sert à rien... si ce n'est risquer de diminuer la qualité du service que lui fournit le site...

Mais cela montre bien que le Referer est une information qui ne doit pas être considérée comme exacte, et que le site ne doit pas se baser dessus pour une fonctionnalité indispensable ou critique.

En conclusion de cet article, on peut dire que certes, le Referer peut parfois poser un risque pour la sécurité... mais que ce risque est largement minimisé par les mesures de sécurité intégrées aux navigateurs les plus répandus...

Et puis, si vous craignez pour votre sécurité et votre vie privée au point de vouloir désactiver ou fausser le Referer... il n'y a plus qu'une chose à faire : déconnectez-vous d'Internet, éteignez votre ordinateur, ne regardez plus la télévision, ne téléphonez plus, et enfermez-vous à double tour dans votre cave !

Après tout, le Referer est avant tout là pour rendre service au visiteur d'un site, que ce soit directement, ou indirectement, par l'intermédiaire du travail que le webmaster peut faire à partir des informations que celui-ci lui apporte !