Le cloaking : spam ou pas spam ? (1ère partie)

Retour au sommaire de la lettre

La notion de "cloaking" divise en ce moment bon nombre de webmasters et de moteurs de recherche sur le Web. C'est un sujet qui passionne notamment les utilisateurs des forums de discussion. Est-ce du spam ou non ? Les moteurs de recherche n'utilisent-ils pas, eux-mêmes, des techniques qui s'en rapprochent ? N'est-ce pas une possibilité technique qui permet, justement, de pallier les insuffisances techniques de certains moteurs ? Les avis sont partagés...

Avant d'aller plus loin, rappelons, si besoin est, ce qu'est le cloaking (et ce qu'il n'est pas...). Il s'agit d'une technique qui permet à un site web de fournir des pages différentes à un internaute et au "spider" (agent, robot) d'un moteur. Prenons un exemple d'une procédure de cloaking :

Etape 1 : Un script (il en existe de nombreux, dans plusieurs langages informatiques, aisément trouvables sur le Web) est installé sur le serveur d'un site web. Ce script est prévu pour se déclencher dès qu'une nouvelle visite survient sur le site web en question.

Etape 2 : lorsqu'une nouvelle visite est détectée, le script tente de détecter s'il s'agit d'un internaute "lambda" ou du "spider" d'un moteur de recherche. Pour cela, il dispose d'une table des "user agents" (les "noms" des spiders : "Scooter" pour AltaVista, "Slurp" pour Inktomi, "Googlebot" pour Google, etc.) ou des adresses IP (dans ce cas, on parle d'"IP Cloaking" ou d'"IP Delivery") des différents robots des moteurs majeurs. Ces informations, là aussi, se trouvent sans trop de difficultés sur le Web. Exemple pour Google :
http://www.searchengineworld.com/spiders/ip_addresses/google.htm
La difficulté essentielle consiste surtout à utiliser des données à jour, celles-ci variant parfois fortement avec le temps... Mais la simple analyse des logs d'un site fournit déjà beaucoup d'informations...

Etape 3 : Si la visite vient d'un visiteur "lambda", la page "normale", "classique", est affichée. Pas de problème dans ce cas.

Etape 4 : Si c'est un robot qui est détecté, le script lui fournit alors une page sépcifiquement écrite pour lui. Selon la façon dont le webmaster a programmé le script, ce dernier fournira toujours la même page, quel que soit le robot, mais il pourra également envoyer une page spécifique pour AltaVista si c'est Scooter qui est détecté, une page pour Inktomi si c'est Slurp, etc.
A noter que certaines personnes font la différence entre "IP Cloaking" et IP Delivery" : l'"IP Cloaking" définira alors une technique permettant de fournir des contenus différents aux humains d'un côté et aux robots des moteurs de l'autre. L'"IP Delivery" sera une technique permettant de fournir des contenus différents en fonction des adresses IP, sans tenir compte des moteurs : par exemple la géolocalisation des internautes (voir plus loin dans cet article).

Le cloaking est donc une technique permettant de fournir des pages web différentes selon que ce soit un internaute ou un moteur qui accède au site web.

Des besoins très variés

Mais quel type de page est alors fourni aux différents moteurs ? A quoi ça sert ? C'est ce point qui divise actuellement webmasters et outils de recherche. En effet, les techniques de cloaking sont utilisées pour de nombreux besoins :

- Référencement d'un site contenant des contraintes techniques à un bon référencement, notamment du Flash ou des pages dynamiques. Si votre site est entièrement créé en Flash, il ne sera que peu visible sur bon nombre de moteurs (voir http://www.flash-moteurs.com/). Il en est de même pour un site dynamique, proposant des adresses (urls) contenant notament des signes "?" (caractères dits "exotiques"). Une solution serait donc de fournir les pages dynamiques ou Flash aux utilisateurs et une version statique et HTML aux moteurs. Ainsi, chacun reçoit une version qu'il peut lire et interpréter à sa guise.

- Référencement de pages optimisées (page alias ou satellites, voir http://docs.abondance.com/question9.html), si la société chargée de référencer un site web ne peut pas toucher au contenu (titre, texte) des pages de son client, ce qui arrive malheureusement assez souvent.

- Possibilité de cacher un code source optimisé par un référenceur professionnel afin que ses concurrents ne s'en inspirent pas :-). Ceci dit, comme nous le verrons plus tard dans la suite de cet article (notamment le mois prochain), il n'est pas très compliqué de "se faire passer" pour un robot afin de lire le dit code optimisé...

- Dérive de la possibilité précedente : certains webmasters peu scrupuleux ont, par le passé, "volé" le code source d'une page web bien positionnée (donc bien optimisée), l'ont réinstallé sur leur site et l'ont fourni à nouveau aux moteurs via des techniques de cloaking, pour être aussi bien classé que la page ainsi spoliée qui ne leur appartenait pas au départ.

- Référencement de pages web sur des expressions n'ayant qu'un lointain rapport avec le contenu de la page, voire sur des noms de concurrents (pour bien positionner le site de Peugeot sur le mot clé "Renault", par exemple, et ceci n'est qu'un exemple, monsieur le webmaster des sites de Peugeot :-)), etc. Dans ces derniers cas, il s'agit très clairement de tentatives de spam.

D'autres motivations pourraient bien entendu être ici listées pour mettre en place des techniques de cloaking.

Qu'est-ce qui n'est pas du cloaking ?

Si nous avons essayé de lister les différents besoins qui peuvent amener un webmaster à faire du cloaking, il convient également de préciser que certaines techniques, qui peuvent s'en rapprocher, n'en sont pas vraiment :

- Tester quel navigateur utilise l'internaute pour lui envoyer soit une page optimisée pour Internet Explorer, soit une page optimisée pour Netscape Navigator, par exemple. Cette technique n'est aujourd'hui pas considérée comme du cloaking, cependant, elle peut être dangereuse car elle peut parfois être assimilée à du spam par les moteurs (voir plus loin).

- Géolocalisation : regarder la langue du navigateur de l'internaute ou effectuer un traitement de son adresse IP pour tenter de savoir dans quel pays il se trouve et lui fournir, en conséquence, un contenu adapté sous plusieurs formes possibles : Soit une redirection vers le site dans la langue de l'internaute, soit des informations différentes. Les outils de recherche sont, d'ailleurs, les premiers à effectuer ce type de "manipulations" : si vous tapez "http://www.altavista.com", il y a de fortes chances que vous soyez redirigé vers "http://fr.altavista.com/", la version française du moteur. Si vous tapez une requête sur le site américain de Yahoo!, vous ne verrez pas apparaître les liens sponsorisés d'Overture si vous vous situez en France. Ces liens publicitaires sont réservés aux internautes géographiquement situés aux Etats-Unis. Alors, les moteurs de recherche font-ils eux-même du cloaking ? Non, la géolocalisation des visiteurs n'est pas, non plus, considérée ainsi.

- Certaines personnes estiment que les offres de référencement payant proposées par AltaVista, Fast ou autres Inktomi sont du cloaking. Ces "trusted feeds" au format XML n'en sont clairement pas, dans le sens où les outils de recherche l'entendent en tout cas...

En fait, définir le cloaking est surtout (comme souvent) une question de bon sens : **on caractérise comme cloaking toute technique visant, en fournissant des pages différentes aux utilisateurs d'un site et aux robots des moteurs, à tromper l'algorithme de recherche de ces moteurs en lui fournissant du "texte caché".**

Spam ou pas spam ?

Une question judicieuse se pose alors : le cloaking est-il du spam ou non ? Car on serait tenté de dire que, tant qu'il pallie les insuffisances, les carences des moteurs (pas de possibilité d'indexer un site en Flash ou dynamique, par exemple), il s'agit d'une bonne chose. Mais que son utilisation à des fins frauduleuses (copies de pages optimisées ou autres) est strictement considérée comme du spam. Comme toujours sur l'Internet, c'est l'utilisation de l'outil qui devrait être jugée et non pas l'outil lui-même... On connaît le vieil exemple : avec un marteau, on peut réaliser un meuble superbe ou taper sur la tête de son voisin. Mais, dans ce cas, qui ira au tribunal ? Le marteau ou celui qui l'a utilisé ?

Selon la classification du spam proposée par Alen Perkins, l'un des spécialistes du domaine aux Etats-Unis (http://www.ebrandmanagement.com/whitepapers/spam-classification/), il n'en est rien : l'usage du cloaking est considéré comme du spam, un point c'est tout, quel que soit la volonté initiale du webmaster. On peut le regretter.

Peut-être serait-il intéressant que les moteurs de recherche créent un label de qualité et permettent ainsi à certains webmasters ou certaines sociétés de référencement, dûment labellisées comme "éthiquement irréprochables", de pratiquer le cloaking uniquement pour des sites présentant des obstacles techniques à un bon référcencement. Vous nous direz, on n'est pas très éloigné, dans ce cas, des offres de "feeds XML" proposés par AltaVista, Fast et Inktomi, entre autres, dans leurs offres de référencement payant comme indiqué précédemment dans cet article. Certes, mais Google ne propose pas de telles offres payantes. Une piste de réflexion pour le moteur de recherche leader ? En effet, de nombreux référenceurs n'ont que très modérement apprécié la mise en ligne d'une page, sur le site de Google, fustigeant leur métier (http://www.google.com/webmasters/seo.html). La solution ne serait-elle pas plutôt de clarifier les choses en créant un label de qualité permettant aux sociétés de référencement de bien faire leur métier sous contrôle du moteur ?

Qu'en pensent les moteurs ?

Puisque nous en parlons, que pensent les moteurs de recherche du cloaking ? Nous leur avons posé la question. Faisons la liste des outils de recherche majeurs dans les mondes anglophone et francophone :

Google parle du cloaking dans sa FAQ aux webmasters (http://www.google.com/webmasters/faq.html#cloaking) : "*The term "cloaking" is used to describe a website that returns altered webpages to search engines crawling the site. In other words, the webserver is programmed to return different content to Google than it returns to regular users, usually in an attempt to distort search engine rankings. This can mislead users about what they'll find when they click on a search result. To preserve the accuracy and quality of our search results, Google may permanently ban from our index any sites or site authors that engage in cloaking to distort their search rankings.*"
La définition de Google est ambigüe : si on la lit de façon exacte, le cloaking réalisée à des fins de spam est interdit, mais la porte semblerait ouverte pour du cloaking mis en place, par exemple, pour des sites web en Flash ou dynamiques... C'est la question que nous avons posé à Google. Voici leur réponse (en V.O.) : "*To clarify, all cloaking is outside of our guidelines and may be penalized*". Cela au moins le mérite de la clarté : toute manoeuvre de cloaking est interdite par le moteur !
On pourrait d'ailleurs penser que Google détient une fonctionnalité assez redoutable pour les webmasters qui désireraient spammer l'index du moteur en faisant du cloaking : la fonction [Cache]. Car, dans ce cas, c'est la page aspirée par le robot, donc le document "cloaké" qui serait visible dans le cache du moteur. Pourtant, bizarrement, lorsqu'un site utilise le cloaking, la page indexée ne semble pas toujours visible via le cache. Prenons un exemple : tapez le mot clé "assurance" sur Google.fr avec l'option "pages francophones" :
http://www.google.fr/search?q=assurance&ie=ISO-8859-1&hl=fr&btnG=Recherche+Google&meta=lr%3Dlang_fr
Sur la première page de résultats apparaît le site de Direct Asurance (en deuxième position sur cette copie d'écran) :

Nous avons pu vérifier, grâce à des outils que nous avons développés (voir plus loin dans cet article), que le site Direct Assurance fait du cloaking pour proposer des pages optimisées aux moteurs de recherche dont Google (vu les résultats fournis par nos outils, il serait difficile pour le webmaster du site de dire le contraire, bien que les pages en question ne semblent pas contenir ce que l'on peut raisonnablement appeler du spam). D'ailleurs, le titre et le texte affichés dans les résultats de Google n'ont rien à voir avec les informations proposées sur le site, si l'on clique sur le lien proposé. Bizarrement, le lien [En cache] n'apparaît pas en face de l'url du site dans les résultats fournis par Google. Etonnant. On ne peut donc pas visualiser la page cloakée.
Explication : le webmaster du site, pas bête ;-), a inséré dans la page web fournie aux moteurs la mention :

<meta name="Robots" content="index, follow, noarchive">

fonctionnalité proposée par Google et qui permet de ne pas voir sa page archivée dans le cache du moteur, ou en tout cas pas visualisée par celle-ci (voir http://www.google.com/webmasters/3.html#B2)... Et le tour est joué...
La fonction de cache n'est donc pas l'arme absolue pour dénicher les tentatives de cloaking sur Google.

AltaVista parle également du cloaking dans ses pages d'aide (http://fr.altavista.com/help/search/faq_web#12) :

> ***11. Does your engine detect and penalize for word stacking, cloaking, multiple page submissions, and other kinds of spamming?***
> *Our anti-spamming techniques are designed to find sites or subnets that submit large numbers of pages with essentially the same content, or that lead to the same content. A Web site that attempts to manipulate search results may be blocked from the AltaVista index. [...]*
> *Here are some specific examples of manipulation that may cause us to block a site from our index:*
> **Pages with text that is not easily read, either because it is too small or is obscured by the background of the page,*
> **Pages with off-topic or excessive keywords,*
> **Duplication of content, either by excessive submission of the same page, submitting the same pages from multiple domains, or submitting the same content from multiple hosts,*
> **Machine-generated pages with minimal or no content, whose sole purpose is to get a user to click to another page,*
> **Pages that contain only links to other pages,*
> **Pages whose primary intent is to redirect users to another page.*
> *Attempts to fill AltaVista's index with misleading or promotional pages lower the value of the index for everyone. We do not allow URL submissions from customers who spam the index and will exclude all such pages from the index.*
> *Please report spam via our e-mail form; please select "Spam Reporting" in the Subject pull-down menu.*

Là aussi, la définition n'est pas claire à 100% et se rapproche de celle de Google. Nous avons donc demandé des éclaircissements à AltaVista. Voici ce qui nous a été répondu : "*Après vérification avec nos ingénieurs, il n'y a pas véritablement de position officielle d'AltaVista au sujet du cloaking, ni de "guideline" que nous publions dans ce cas. Notre attitude est assez semblable à celle affichée sur le site de Google sur ce sujet, et si nous trouvons des pages qui spamment le moteur par ce biais, nous les banissons de l'index.*"

Inktomi propose quelques informations sur son site web (http://www.inktomi.com/products/web_search/spampolicyfaq.html) :

> ***Q: Does Inktomi allow "cloaking"?***
> *A: Inktomi treats every document in the index equally whether the documents are crawled or submitted directly to the index - this includes HTML pages, XML feeds and Inktomi Data Interchange Format (IDIF) documents. If a document is found to be deceptive or misleading, it will be removed.*
> ***Q: What if I present one page to Internet Explorer users, and a different page to Netscape users?***
> *A: That's fine. If the purpose is to serve alternate pages to different human users, based on locality, browser, machine type etc., we do not consider that cloaking.*

Idem : la réponse n'est pas très claire. Nous n'avons malheureusement pas pu joindre de représentants d'Inktomi qui puissent nous en dire plus à ce sujet, sur lequel la position du moteur pourrait éventuellement changer prochainement suite à son rachat par Yahoo!.

Fast, par l"intermédiaire de son moteur AllTheWeb, propose une page sur la spam (http://www.alltheweb.com/info/about/spam_policy.html) mais n'y parle pas de cloaking. Nous avons donc posé la question à Fast. Leur réponse est claire : "*Nous n'acceptons pas le cloaking. La réponse est brève mais éloquente*". Pas de place pour le doute, donc, chez Fast :-)

Pas d'informations sur le cloaking sur le site de Voila (http://www.voila.fr/). Nous nous sommes donc adressé à Pierre Aubert, responsable technique du moteur : "*La politique de Voila est assez claire. Le cloaking est presque systématiquement pris pour du spam et entraine l'élimination de la base. Par contre, comme on ne fait pas de recherche systématique de cloaking, pas mal de gens passent à travers. Ca coûte cher et de toutes facons, ça n'est pas très efficace, nos IPs étant connues comme le loup blanc. On a cependant autorisé certains gros sites de commerce électronique, a priori des gens qui ont pris contact avec nous précédemment, à nous renvoyer un html special, détecté via notre "User Agent", VoilaBot. Je pense que l'on va généraliser ça, en tout cas, pour les sites pros prêts à signer un contrat de bon fonctionnement avec nous. Les résultats sont évidement positifs pour ces sites comme pour nous. Pour eux, seul le contenu qui les intéresse est pris en compte et de notre côté, on y gagne aussi avec des pages plus légères, sans pubs... La détection de langues et de thématiques marchent mieux aussi sur ce genre de pages.*"

Sébastien Richard, d'Exalead (http://www.exalead.com/), nous donne son opinion sur la question : "*Chez Exalead, nous souhaitons indexer des documents les plus proches de ce que le surfeur obtiendra lors de son arrivée sur la page web. Dans cet ordre d'idée, nous n'indexons pas les balises META. Mais, nous sommes aussi conscients de certaines limites techniques de notre spider (notamment sur l'indexation des sites en Flash) et c'est pourquoi nous tolérons le cloaking dans la limite où le texte renvoyé au moteur est en accord avec la version "browsable" du site.*

Nous gardons le secret sur les méthodes que nous utilisons pour vérifier l'adéquation du contenu proposé au moteur avec la version "browsable" mais dans le cas où cette adéquation ne serait pas vérifiée, des mesures sont prises à l'encontre du site coupable pouvant aller jusqu'à son bannissement."

Fabien Menemenlis, l'un des concepteurs du moteur Dir.com (http://www.dir.com/), répond à la question du cloaking : "*Il est difficile de faire une analyse complète d'une page, notamment à cause du JavaScript très difficilement interprétable par un robot. Tant que l'information reste pertinente sur le contenu du site, qu'elle est de taille raisonnable, il n'y a pas de raison de bannir le site. Par contre, les sites qui utilisent ces méthodes pour rediriger vers des sites à contenu payant ou qui trompent visiblement l'utilisateur doivent être bannis : ces sites combinent bien souvent d'autres méthodes facilement détectables, notamment au niveau de la "richesse" du texte qui les placent dans la catégorie de spammeurs, et se retrouvent donc assez naturellement mis en liste noire.*"

Christelle Ott, d'Antidot (http://www.antidot.net/), répond à notre question au sujet du cloaking : "*Nous n'avons rien contre le principe du cloaking, et nous le trouvons même utile dans certains cas (accès à une base de données dynamique, meilleure analyse du contenu du document, etc). Cependant, tout comme celle des balises META, cette technologie peut être utilisée pour tromper un moteur (et ses utilisateurs) et donc pour tenter de remonter des sites à partir de mots clés qui ne reflètent pas leurs contenus.*
AFS permet de visualiser la page telle que ses robots l'ont téléchargé (et donc le cas échéant la version "cloakée") - via le lien "archive" présent à côté de chaque réponse. Les pages "cloakées" sont acceptées sans problème dans notre index si elles permettent de mieux faire indexer leur contenu. Si par contre nous détectons une tentative de tromperie, alors le site complet est mis sur nos listes noires. La méthode utilisée par google est propriétaire (l'attribut "noarchive" n'est pas standard). Pour ne pas avoir ses pages archivées par AFS il faut en faire la demande explicite à support@antidot.net, ce qui nous permet de filtrer les demandes entrantes."

Gilbert Wayenborgh, DeepIndex (http://www.deepindex.com/), prône, quant à lui, la "tolérance technique" : "*En effet, lorsque Deepindex a été créé, l'approche a été certes grand public, mais avec une franche collaboration "référencement". Si un certain nombre de sites utilisent cette technique (parmi tant d'autres) pour des raisons techniques (Flash, Javascript, etc..), certains l'utilisent à des fins de spam. C'est dans ce dernier cas que nous le considérons comme "interdit" et risquons d'éliminer le "domaine" entier. Deepindex a une vision "industrielle" et "marché". Dans cette vision industrielle il y a plusieurs mondes qui se cotoyent : les outils de recherche, les référenceurs, et le public bien entendu. Les besoins des uns et des autres doivent donc être satisfaits.*"

Comment les moteurs combattent-ils le cloaking ?

Nous avons passé en revue la définition du cloaking, ce qu'il est et ce qu'il n'est pas. Nous avons entrevu l'opinion, parfois fort tranchée, parfois plus nuancée, des moteurs de recherche à ce sujet.

Mais comment ces moteurs se prémunissent-ils contre toute tentative de spam utilisant le cloaking ? Comment détectent-ils les fraudes ? Comment combattent-ils ce "fléau" ?

En fait, il semble exister deux manières distinctes de le faire :

- La première consiste à changer très souvent les "user-agents" et / ou adresses IP des robots utilisés. De ce fait, les bases de données utilisées par les spammeurs ne sont pas à jour et les robots "passent au travers" des filtres mis en place par les webmasters. La première fois, en tout cas, puisque le spider laissera sa trace dans les logs du serveur et ses nouvelles coordonnées risquent fort d'être insérées dans la base de données à sa prochaine venue. Difficile d'être discret... Tout cela ressemble fort à une version électronique des gendarmes et des voleurs (il en est hélas souvent ainsi pour les techniques de spam).

- La seconde consiste à mettre en place plusieurs robots : l'un s'identifie comme le spider du moteur et l'autre se fait passer pour un navigateur "lambda" (par exemple Explorer ou Netscape). Le moteur compare alors les résultats fournis par le site aux deux robots : si les différences sont notables, il se peut qu'une procédure de vérification approfondie soit mise en place ou que la page (voire le site entier) soit tout simplement bannie de l'index (procédure de liste noire).

Ces principes ont été mis en place sur un outil que nous avons développé, au niveau du Réseau Abondance, et qui nous sert à détecter les tentatives de cloaking sur les moteurs de recherche. Il nous permet également de faire des tests pour savoir quel moteur de recherche est le plus "sensible" à ce phénomène. Mais tout cela fera partie de la deuxième de cette article qui paraîtra le mois prochain. Patience...

Pourquoi est-ce dangereux ?

Le cloaking et sa détection par certains moteurs majeurs peuvent avoir un côté dangereux. En effet, imaginons que vous mettiez en place un système qui suive la procédure suivante :
- Dès qu'un visiteur arrive sur votre site, vous testez quel navigateur il utilise.
- 4 cas se présentent : Internet Explorer, Netscape Navigator, Opera ou un autre navigateur. Si l'un des trois premiers cas survient, vous proposez une page spécifiquement optimisée pour le logiciel en question. Pas de problèmes jusque là... Si vous ne détectez aucun de ces 3 logiciels, vous estimez alors qu'il s'agit d'un navigateur textuel (de type Lynx) et vous fournissez une version, elle aussi textuelle, de la page.

Dans ce cas, si le moteur effectue des vérifications en comparant les différentes versions (voir paragraphe précédent : *comment les moteurs combattent-ils le cloaking ?*), la version textuelle de votre page risque fort d'être prise pour une page cloakée spécifiquement écrite pour les moteurs, donc d'être prise pour du spam. Et votre site risque la liste noire...

Attention, donc, si vous mettez en place des scripts qui effectuent des tests "discriminatifs", quels qu'ils soient, lors de l'arrivée d'un internaute : faites bien attention à ce que vos intentions, même si elles sont empreintes de bonne foi, ne soient pas mal interprétées par les moteurs ! Vérifiez bien la façon dont vous mettez en place d'éventuels filtres de comparaisons. Attention à ne pas être le Monsieur Jourdain du spam, tenant une "fraude au moteurs" sans le savoir...

Nous avons vu dans cet article que certains moteurs ont une opinion assez nuancée sur le sujet du cloaking. On peut résumer cela ainsi : "oui au cloaking tant que ces techniques ne sont pas utilisées pour spammer". Vous savez donc ce qui vous reste à faire... Dans tous les cas, si vous désirez mettre en place des procédures de cloaking (même de façon tout à fait honnête), réflechissez-y bien à deux fois, car vous entrez dans une compétition qui se jour sur "terrain glissant", et donc à hauts risques pour vos sources d'information...

Terminons cet article en indiquant que le cloaking est surtout utilisé par les webmasters pour influencer les documents fournis aux robots des moteurs. Les annuaires ne sont pas touchés par ces problèmes. Cependant, il semblerait tout à fait possible, techniquement parlant, de fournir aux documentalistes de certains annuaires (qui disposent parfois d'un numéro IP spécifique) une version du site "propre" lorsque ces derniers viennent le visiter pour l'évaluer, mais d'en proposer une autre aux internautes "lambda". On n'en est pas encore là sur le Web. Ouf car, là encore, les risques de dérives sont importants...

Le mois prochain : présentation d'un utilitaire "détecteur de cloaking" que nous avons développé et qui permet d'indiquer quels moteurs de recherche sont le plus sensibles au phénomène.

Terminons notre article de ce mois avec ***quelques liens intéressants sur le sujet*** :

Promo-Web
http://www.promo-web.org/Optimisation/cloaking.htm

Why cloaking is always a bad idea
http://www.webyield.net/newsletters/009.htm

Ending the debate over cloaking
http://searchenginewatch.com/sereport/03/02-cloaking.html

To clak or not to cloak?
http://searchenginewatch.com/searchday/01/sd0913-ses-cloaking.html

Et un grand Merci à Jean-Philippe Caste et Mark Barlow (Altavista), Sébastien Richard (Exalead), Pascal Gayat (Fast), Debbie Frost (Google), Pierre Aubert et Thierry Pourvendier (Voila), Fabien Menemenlis (Dir.com), Christelle Ott (Antidot) et Gilbert Wayenborgh (DeepIndex) d'avoir bien voulu répondre à nos questions.