

**Le cloaking : spam ou pas spam ? (2ème partie)**

[Retour au sommaire de la lettre](#)

Le mois dernier, nous avons tenté de définir ce qu'était (et ce que n'était pas) le cloaking. Nous avons également fourni l'avis de chaque moteur de recherche sur ces techniques. Ce mois-ci, nous allons tenter de voir quels sont les moteurs de recherche qui sont sensibles au cloaking et si, tout simplement, cette technique est très répandue ou non sur le Web. Quels sont les moteurs qui contiennent des résultats "cloakés" dans leurs listes de liens ? Ces derniers sont-ils légion ou apparaissent-ils de façon anecdotique dans les résultats ? C'est ce que nous allons essayer de savoir...

Pour mener à bien cette étude, nous avons établi la méthodologie suivante :

- Nous avons pris en compte des 100 mots clés et expressions le plus souvent demandés sur le Web francophone, selon les données fournies depuis peu par le Lycos 100 (<http://www.recherche.lycos.fr/lycos100/>).

- Nous avons développé (merci à la société Brioude Internet !) un outil qui effectue de façon automatique la procédure suivante pour Google, AltaVista, Voila et AllTheWeb :
  - \* Chaque mot clé du Lycos 100 est soumis au moteur.
  - \* Pour chaque requête soumise, les 10 liens de résultats (classés de 1 à 10, donc) proposés par le moteur sont "capturés".
  - \* Chacun des liens est "cliqué" par logiciel suivant trois accès différent : en se faisant passer pour un navigateur Explorer, un navigateur Netscape et le robot du moteur en question.
  - \* Dans un premier temps, la taille des fichiers renvoyés est comparée : si elle est équivalente dans les 3 cas, cela signifie qu'il y a de très forte chance qu'aucune distinction ne soit faite entre les navigateurs d'un côté et le robot du moteur de l'autre, donc que le site en question ne fait pas de cloaking.
  - \* Si la taille des fichiers est différente, les 3 codes HTML correspondants sont retournés par notre applicatif et vérifiés manuellement pour savoir s'il y a cloaking ou non.

Voici les résultats que nous avons obtenus pour les 4 moteurs testés (**NTCD** = Nombre de Tentatives de Cloaking Détectées) :



<b>Mot clé</b>	<b>NTCD</b>	<b>Mot clé</b>	<b>NTCD</b>	<b>Mot clé</b>	<b>NTCD</b>	<b>Mot clé</b>	<b>NTCD</b>
sexe	0	mp3	0	sport	0	jenifer	0
google	0	alizee	0	sex	0	traduction	1
caramail	0	emule	0	gîtes de France	0	charme	0
voissa	0	emploi	0	telechargercom	0	tarot de marseille	0
chat	0	gay	0	annuaire inversé	0	wallpaper	0
pages jaunes	0	annuaire	0	lingerie	0	code de la route	0
horoscope	0	fnac	0	sexyloo	0	corse	0
ANPE	0	lettre de motivation	0	loto	0	jeux vidéo	0
sncf	0	tf1	0	traducteur	0	moto	0
fonds écran	0	sms gratuit	0	rencontre	0	mariage	0
sexe gratuit	0	itinéraire	0	webcam	0	air france	0
immobilier	0	eminem	0	argus	0	photos	0
clipart	0	string	0	cinema	0	aol	0
kazaa	0	michelin	0	tatouage	0	jennifer lopez	0
meteo	0	paroles de chansons	0	impot	0	météo neige	0
revebebe	0	hotmail	0	irak	0	warez	0
humour	0	boursorama	0	manga	0	blagues	0
ip-label	0	wanadoo	0	hollywood	0	smallville	0
clara morgane	0	voila	0	genealogie	0	msn	0

divx	0	ecran de veille	0	worldspace	0	avril lavigne	0
test	0	telecharger	0	gratuit	0	sims	0
yahoo	0	sfr	0	tuning	0	email	0
lorie	0	dictionnaire	0	edonkey	0	mappy	0
hotel	0	abcoeur	0	harry potter	0	cartes virtuelles	0
jeux	0	musique	0	moteur de recherche	0	crumble	0

**Résultat** : 1 présomption de cloaking sur Google.

A noter un superbe cloaking (la plus forte tentative de notre test, tous outils de recherche réunis, une belle et grande tentative de spam) pour le mot clé "traduction" de la part... d'un autre outil de recherche (qui ne se situe pas parmi le "Top 10", mais pas si loin)... Eh oui...



<b>Mot clé</b>	<b>NTCD</b>	<b>Mot clé</b>	<b>NTCD</b>	<b>Mot clé</b>	<b>NTCD</b>	<b>Mot clé</b>	<b>NTCD</b>
sexe	0	mp3	0	sport	0	jenifer	0
google	0	alizee	0	sex	0	traduction	0
caramail	0	emule	0	gîtes de France	0	charme	0
voissa	0	emploi	0	telechargercom	0	tarot de marseille	0
chat	0	gay	0	annuaire inversé	0	wallpaper	0
pages jaunes	0	annuaire	0	lingerie	0	code de la route	0
horoscope	0	fnac	0	sexyloo	0	corse	0
ANPE	0	lettre de motivation	0	loto	0	jeux vidéo	0
sncf	1	tf1	0	traducteur	0	moto	0
fonds écran	0	sms gratuit	0	rencontre	0	mariage	0
sexe gratuit	0	itinéraire	0	webcam	0	air france	0
immobilier	0	eminem	0	argus	0	photos	0
clipart	0	string	1	cinema	0	aol	0
kazaa	0	micelin	0	tatouage	0	jennifer lopez	0
meteo	0	paroles de chansons	1	impot	0	météo neige	0
revebebe	0	hotmail	0	irak	0	warez	0
humour	0	boursorama	0	manga	0	blagues	0
ip-label	0	wanadoo	0	hollywood	0	smallville	0
clara morgane	0	voila	0	genealogie	0	msn	0
divx	0	ecran de veille	0	worldspace	0	avril lavigne	0
test	0	telecharger	0	gratuit	0	sims	0
yahoo	0	sfr	0	tuning	0	email	0
lorie	0	dictionnaire	0	edonkey	0	mappy	0
hotel	0	abcoeur	0	harry potter	0	cartes virtuelles	0
jeux	0	musique	0	moteur de recherche	0	crumble	0

**Résultat** : 3 présomptions de cloaking sur Voila.



<b>Mot clé</b>	<b>NTCD</b>	<b>Mot clé</b>	<b>NTCD</b>	<b>Mot clé</b>	<b>NTCD</b>	<b>Mot clé</b>	<b>NTCD</b>
sexe	0	mp3	0	sport	0	jenifer	0
google	0	alinee	0	sex	0	traduction	0
caramail	0	emule	0	gîtes de France	0	charme	0
voissa	0	emploi	0	telechargercom	0	tarot de marseille	0
chat	0	gay	0	annuaire inversé	0	wallpaper	0
pages jaunes	0	annuaire	0	lingerie	0	code de la route	0
horoscope	0	fnac	0	sexyloo	0	corse	0
ANPE	0	lettre de motivation	0	loto	0	jeux vidéo	0
sncf	0	tf1	0	traducteur	0	moto	0
fonds écran	0	sms gratuit	0	rencontre	0	mariage	0
sexe gratuit	0	itinéraire	0	webcam	0	air france	0
immobilier	0	eminem	0	argus	0	photos	0
clipart	0	string	0	cinema	0	aol	0
kazaa	0	micelin	0	tatouage	0	jennifer lopez	0
meteo	0	paroles de chansons	0	impot	0	météo neige	0
revebebe	0	hotmail	0	irak	0	warez	0
humour	0	boursorama	0	manga	0	blagues	0
ip-label	0	wanadoo	0	hollywood	0	smallville	0
clara morgane	0	voila	0	genealogie	0	msn	0
divx	0	ecran de veille	0	worldspace	0	avril lavigne	0
test	0	telecharger	0	gratuit	0	sims	0
yahoo	0	sfr	0	tuning	0	email	0
lorie	0	dictionnaire	0	edonkey	0	mappy	0
hotel	0	abcoeur	0	harry potter	0	cartes virtuelles	0
jeux	0	musique	0	moteur de recherche	0	crumble	0

Résultat : 0 présomption de cloaking sur AltaVista.



<b>Mot clé</b>	<b>NTCD</b>	<b>Mot clé</b>	<b>NTCD</b>	<b>Mot clé</b>	<b>NTCD</b>	<b>Mot clé</b>	<b>NTCD</b>
sexe	0	mp3	0	sport	0	jenifer	0
google	0	alinee	0	sex	0	traduction	0
caramail	0	emule	0	gîtes de France	0	charme	0
voissa	0	emploi	0	telechargercom	0	tarot de marseille	0
chat	0	gay	0	annuaire inversé	0	wallpaper	0
pages jaunes	0	annuaire	0	lingerie	0	code de la route	0
horoscope	0	fnac	0	sexyloo	0	corse	0
ANPE	0	lettre de motivation	0	loto	0	jeux vidéo	0
sncf	0	tf1	0	traducteur	1	moto	0
fonds écran	0	sms gratuit	0	rencontre	0	mariage	0
sexe gratuit	0	itinéraire	0	webcam	0	air france	0
immobilier	0	eminem	0	argus	0	photos	0
clipart	1	string	0	cinema	0	aol	0
kazaa	0	micelin	0	tatouage	0	jennifer lopez	0
meteo	0	paroles de chansons	0	impot	0	météo neige	0
revebebe	1	hotmail	0	irak	0	warez	0

humour	0	boursorama	0	manga	0	blagues	0
ip-label	0	wanadoo	0	hollywood	0	smallville	0
clara morgane	0	voila	0	genealogie	0	msn	0
divx	0	ecran de veille	0	worldspace	0	avril lavigne	0
test	0	telecharger	0	gratuit	0	sims	0
yahoo	0	sfr	0	tuning	0	email	0
lorie	0	dictionnaire	0	edonkey	0	mappy	0
hotel	0	abcoeur	0	harry potter	0	cartes virtuelles	0
jeux	0	musique	0	moteur de recherche	0	crumble	0

**Résultat** : 3 présomptions de cloaking sur AllTheWeb.

Le mot clé "traduction" est l'objet d'un cloaking identique à celui détecté sur Google, de la part du même outil de recherche.

### Conclusion

On peut le voir, les techniques de cloaking sont très peu présentes dans les pages de résultats des moteurs de recherche (7 tentatives détectées sur 400 requêtes, donc 4000 liens testés, soit 0,17% des résultats), notamment pour des mots clés en français (car nous avons la faiblesse de croire que notre test est fiable :-)). Cette faible présence du cloaking peut être due à deux raisons principales :

- La technique n'est pas très usitée par les webmasters francophones.
- La technique est fortement combattue avec succès par les moteurs.

Pour compléter notre propos, il faut indiquer que nous avons identifié de très nombreux cas où les trois tailles de fichier (Explorer, Navigateur, robot) étaient différentes. Mais, dans une immense majorité des cas, il s'agissait simplement de pages légèrement différentes et optimisées pour chaque navigateur plus une troisième version pour "le reste du monde des logiciels de navigation" (Opera, etc.). Les outils de recherche (Yahoo!, Google, Lycos, MSN, etc.) sont les premiers à mettre en place ce type de double ou triple affichage selon le navigateur. Difficile donc de parler de cloaking ici...

D'autre part, le fait de simuler un robot de moteur nous a indiqué qu'un certain nombre de sites "plantent" (message d'erreur) lorsque le robot arrive. Exemple d'un tel message :

```
"Microsoft OLE DB Provider for ODBC Drivers error '80040e14'  
[Microsoft][ODBC SQL Server Driver][SQL Server]Syntaxe incorrecte vers le mot clé 'where'.  
/robot.asp, line 50"
```

Le nom du fichier ("robot.asp") semble peut-être, ici, indiquer une tentative mal réalisée de cloaking...

Autre exemple, sur un des sites de Microsoft :-)

```
"Microsoft VBScript runtime error '800a000d'  
Type mismatch: '[string: "o"]'  
/intl/bver.inc, line 8"
```

Allez vous étonner, ensuite, que le site ne soit pas (ou soit mal) indexé dans les moteurs de recherche...

Donc, selon notre test, très peu de présence de cloaking et donc de spam dans les résultats de moteurs majeurs. Comme ces derniers ont souvent une attitude très ferme contre ces pratiques, tout semble aller pour le mieux dans le meilleur des mondes...