

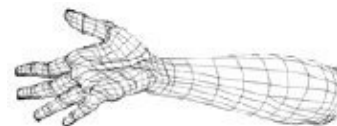


numéro soixante-huit (#68)

le vendredi 20 février 2003



Sommaire



- [Sommaire](#)
- [Editorial](#)
- [Le monde selon Google](#)
GoogleBot expliqué en détails
- [La page des moteurs et métamoteurs](#)
Les métamoteurs – SurfWax
- [La chronique du Web invisible](#)
Le web invisible – les outils de recherche spécialisés
- [Enfin, les informations](#)
- [Les outils de recherche francophones](#)
- [Conclusion](#)

[Le dernier Moteurzine en html](#)

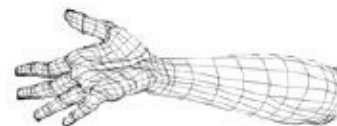
[Le dernier Moteurzine en PDF](#)

[Les archives de Moteurzine](#)

[Ecrire à Moteurzine](#)



Éditorial



Inktomi, Altavista... mais quand cela va donc s'arrêter !

Et oui, c'est le moment... tout le monde se fait racheter... et cela aurait même pu arriver à votre journal préféré sur les outils de recherche, MoteurZine. Lisez donc...

CMGI, en ayant bradé Altavista, s'est retrouvé avec 60 millions de dollars dans les poches ainsi que du papier dont la valeur varie tous les jours. Que faire de tout cet argent si ce n'est le réinvestir ! CMGI a

donc envoyé plusieurs de ses avocats d'affaire négocier âprement le rachat de MoteurZine (oui, oui... ils ne connaissent pas Enfin mais uniquement MoteurZine, chez CMGI). Ils ont proposé 500 000 dollars pour ce journal d'information (d'investigation). Il va de soit que j'ai refusé. Qui veut des dollars, aujourd'hui ??? J'ai donc fait une contre-proposition à 50 000 euro. En entendant cela, les avocats de CMGI sont partis en claquant de la porte, ils n'aiment pas qu'on les prenne pour des c*ns en leur proposant une monnaie de "république bananière".

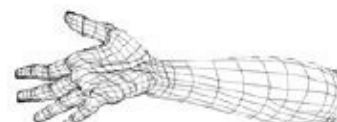
Résultat, aujourd'hui, j'ai gardé mon indépendance éditoriale et mon indépendance d'esprit pour MoteurZine ainsi qu'Enfin (et les autres sites) en travaillant la journée comme plongeur dans une sandwicherie industrielle grecque, ou peut-être turc... enfin, peu importe. Merci à tous.

...et bonne lecture !

par CHRIS HÉDÉ



Le monde selon Google



GoogleBot expliqué en détails

GoogleBot est le nom du robot d'indexation de Google. Ce robot est programmé pour fonctionner sur des centaines de machines à la fois, avec des adresses IP différentes. Il faut dire qu'il a 3 milliards de documents à mettre à jour régulièrement, et des millions de nouveaux à découvrir...

Dans la "famille des GoogleBot" on distingue deux sortes de robots :

- le **Fresh Crawler**, dont l'adresse IP commence par 64.68.82., correspond au robot qui indexe les nouvelles pages trouvées par Google ; une fois visitées par ce robot, les pages apparaissent dans Google seulement quelques jours.
- le **Deep Crawler** (ou **Full Crawler**), dont l'adresse IP commence par 216.239.46., correspond au robot qui effectue une indexation massive de tous les documents connus de Google, en général pendant environ une semaine, juste après la Google Dance.

Le Fresh Crawler n'indexe que les documents aux formats HTML et texte (formats MIME text/html et text/plain), tandis que le Deep Crawler indexe également d'autres types de documents (PDF, PostScript, Word, Excel, PowerPoint...).

Le Deep Crawler a pour objectif de faire une indexation massive de chaque site qu'il visite. Il est difficile de décrire selon quel algorithme il visite les pages, car cela dépend de plusieurs facteurs (liés au site) et du nombre de robots utilisés pour indexer le site. Les principaux critères ayant une influence sur la fréquence et le nombre de visites d'une page sont le PageRank et la fréquence de mise à jour par le webmaster. Il est possible également que la distance (en nombre de liens) de la page d'accueil joue un rôle.

Afin d'éviter une surcharge de votre serveur, GoogleBot espace ses visites dans le temps. D'autre part, il respecte le protocole d'exclusion des robots et commence donc toute indexation par la consultation du **fichier robots.txt** (si vous n'en avez pas, cela génère donc des erreurs 404, donc il vaut mieux en mettre un, même s'il reste vide).

Pour savoir si GoogleBot est venu sur votre site, il vous suffit de consulter vos fichiers log (journal des

requêtes de votre site, stocké sur votre serveur). Si vous n'avez pas accès à ces fichiers, ou si vous ne savez pas comment les utiliser, vous pouvez utiliser **GoogleStats**. Il s'agit d'une application gratuite Open Source écrite en PHP et MySQL, permettant d'analyser en détails les visites de Google sur votre site. A partir de la version 2.0, il sera bientôt possible de gérer autant de robots que vous le voulez !

Pour faciliter l'indexation de votre site, évitez à tout prix de passer des identifiants de session dans vos URL. En effet dans ce cas GoogleBot ne peut jamais terminer l'indexation d'un site, puisqu'il obtient un nouvel identifiant à chaque visite (il "pense" donc trouver une nouvelle page).

Pour les pages dynamiques, il est grandement recommandé d'utiliser la technique de l'**URL rewriting**.

Enfin assurez-vous que votre site est accessible, sinon en cas de visite de GoogleBot pendant une panne, il risque de se "vexer" et de ne plus revenir...

par Olivier DUFFEZ de **WebRankInfo**



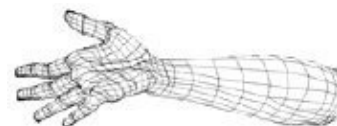
Rechercher sur Enfin



Les pages : françaises anglaises



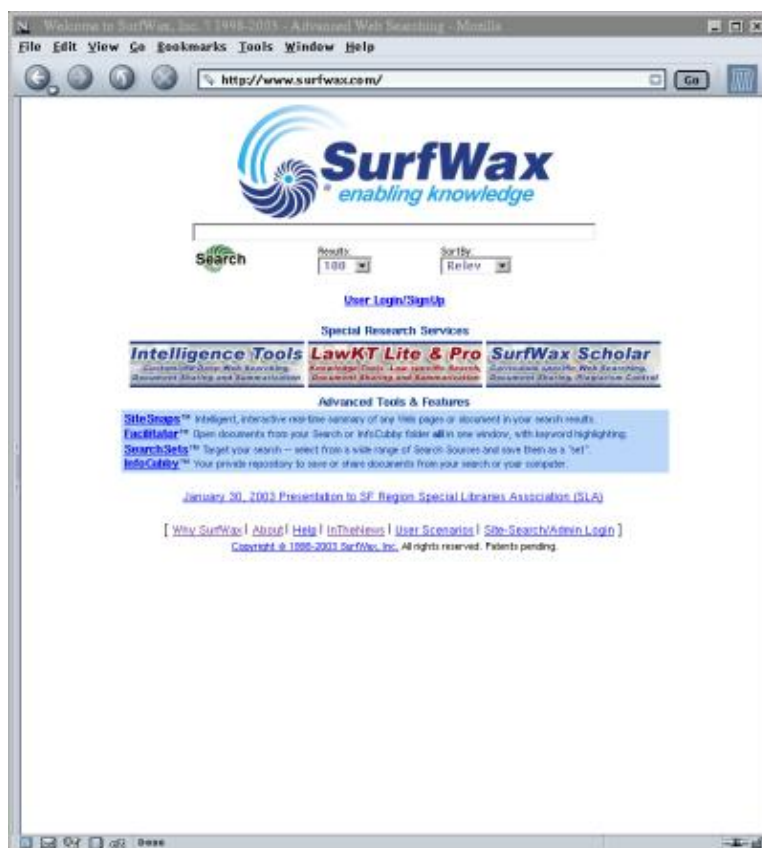
La page des moteurs et métamoteurs



Les métamoteurs – SurfWax

SurfWax ou comment découvrir ce que veut dire "chercher" !

SurfWax, vous connaissez ? Non ??? Quelle erreur !! SurfWax fait parti des meilleurs métamoteurs disponibles sur le Web. Il offre des possibilités inconnues par une grande partie de ses concurrents et surtout, une fois que l'on essaye SurfWax, il est difficile de regarder les autres métamoteurs de la même manière et par conséquent, on revient très rapidement à cette merveille.



Copie écran de la page de garde du métamoteur SurfWax

Avant tout, il est important de signaler que SurfWax est un métamoteur établi et reconnu par les professionnels. Il n'est pas apparu hier pour disparaître demain. SurfWax existe depuis 1998.

Dans sa recherche par défaut, SurfWax trouve l'information dans les sites suivants : Yahoo, Yahoo News, About, AOL, MSN, Dmoz, Encarta, SearchEdu et Looksmart. Ces sites sont tous des références... sauf le dernier dont la présence suprends un peu... mais, ce sont eux les spécialistes !

Lors d'une requête, SurfWax divise la page des résultats en deux. Les liens s'affichent sur la gauche. Sur la droite, nous découvrons au fur et à mesure des informations supplémentaires. Si un lien nous intéresse, en cliquant sur la loupe juste devant, nous voyons apparaître de nombreuses données intéressantes :

- le nom du site
- son url
- le nombre de liens, d'images, de mots et de formulaire sur la page
- la méta description sous le nom "author summary"
- des extraits de phrases reprenant les mots de la requête
- un focus sur les mots ou expressions qui semblent les plus pertinents en fonction de la requête. D'ailleurs, cliquer dessus reformule la question en rajoutant ces éléments pour la recherche et donc donnant une plus grande précision au niveau des résultats.

Et tout cela s'appelle le "SurWax SiteSnaps".



Copie écran de la page résultat du métamoteur SurfWax

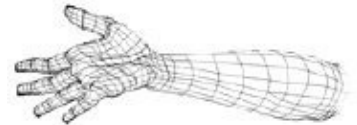
Fini ?... mais non, SurfWax propose d'autres fonctionnalités comme le "SurfWax Facilitor". Cet outil permet de naviguer beaucoup plus facilement sur les résultats. En effet, tous les résultats (enfin une dizaine) s'affichent dans une grande page texte, complètement débarrassée des images et du html offrant ainsi une plus grande facilité pour la lecture et la recherche de l'information.

Le dernier point et le plus intéressant, c'est que SurfWax possède ENORMEMENT de sources d'information et sur de nombreuses thématiques. La version de base (donc gratuite) offre la possibilité de sélectionner 10 sources d'information parmi 500 et pour la version payante, c'est 15 parmi 2000.

Alors, qu'attendez-vous ??? Allez donc essayer de suite SurfWax !



Espace publicitaire



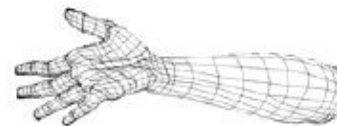
Développez votre indice de popularité...

- pour 0 euros en passant 3 jours à référencer votre site dans les **150** outils généralistes d'Enfin (**ICI**).
- pour 55 euros en passant 5 minutes en commandant un référencement dans les **150** outils généralistes d'Enfin (**LA**!).

– **votre publicité lue par 15.000 personnes !** –



La chronique du Web invisible



Le web invisible – les outils de recherche spécialisés

Flipper

Flipper est un métamoteur spécialisé dans la recherche sur le web invisible à dominante e-commerce et bases de données médias. Lancé en avril 2002 par la société Quigo, spécialiste des techniques de crawling sur le web invisible, Flipper va donc interroger des grandes bases de données, marchandes ou non, de l'univers du cinéma, des livres, des disques, des célébrités (comme Barnes and Noble, Buy.com, Imdb.com, ebooks.com, Moviefone.com...). Flipper interroge également les bases de dépêches d'actualités telles que CNN, New York Times, Advisor, Washington Post, Time, Business 2.0....

La plupart des moteurs de recherche indexent du texte mais sont bloqués lorsqu'ils arrivent sur un formulaire ou une boîte de requête. Quigo a une approche inverse : ses "crawlers" ignorent la plupart des pages web texte et recherchent les pages avec un formulaire de requête.

Il ramène ses pages et des algorithmes décodent ces formulaires pour déterminer quels types de requêtes ils requièrent : mot de passe, nom, chaînes de caractère, variable numérique..Quigo va ensuite créer un agent qui peut spécifiquement interagir avec le formulaire.

IncyWincy, The Invisible Web Search Engine

Incy Wincy de la société LOOP improvements utilise l'Open Directory Project. En crawlant à l'intérieur des sites de cet annuaire, le moteur Incy Wincy construit un index de milliers de moteurs de recherche internes rencontrés. Exemple : Une requête sur la "CIA" permettra de détecter, sur les sites proposés par l'ODP, le moteur de l'United States Intelligence Community. Une recherche sur "Chimie" affichera, parmi les résultats les moteurs internes du Bottin de la Chimie et de l'Ecole Nationale Supérieure de Chimie de Rennes. On peut ensuite réutiliser ces moteurs par un simple click pour préciser la recherche. Une démarche originale.

Turbo10

Turbo10, le métamoteur britannique, utilise des moteurs de recherches spécialisés permettant de rechercher dans des bases de données ou des documents du "web profond" dans les domaines Médicaux (Hardin, Intelihealth, Medscape, HelathCentral...), Sport (Abcsports, Skysports), Actualités (Abcnews, TheTimes, Cnn, BBC, Msnbc, Yahoo News...),

Le module Topic clusters offre une technique de catégorisation qui permet de classer les résultats dans des catégories pertinentes. Ex : la requête " Agriculture" génère les dossiers department, food, trade, forestry, rural...

Thunderstone web site catalog

Thunderstone permet de rechercher des sites via son moteur ou les catégories de son annuaire thématique. Thunderstone privilégie la qualité à la quantité : l'essentiel y est : Politique, Sciences, Economie, Commerce, Santé, Loisirs, Société, Culture, Art, Informatique, Internet, Actualités et Médias...Pour chacune de ces rubriques s'affichent des sous-catégories à la droite de l'écran.

Profusion

Le métamoteur Profusion (d'Intelliseek) propose notamment de rechercher des résultats dans le "web invisible" c'est à dire des bases de données de texte, sons, images, articles, magazines, livres...). Profusion propose de sélectionner les sources d'information et moteurs parmi des thématiques verticales (gouvernement, emploi, science, santé, juridique, enseignement, technologies, voyages,

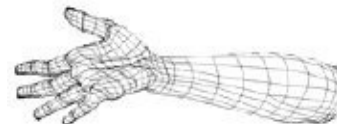
finance, art, business...). Ainsi, par exemple, vous pouvez choisir d'effectuer une recherche au sein d'encyclopédies (Encarta, Encyclopedia.com, InfoPlease, Internet Public Library, References) ou de bases de données et publications dédiées à la biologie (BioLinks, Discover Magazine, NBII ...).

Retrouvez la suite (les répertoires sélectifs et les portails du web invisible) dans le prochain numéro de MoteurZine.

par Christophe ASSELIN de Intelligence-center.com



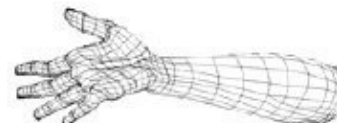
Espace publicitaire



– vosre publicité lue par 15.000 personnes ! –



Enfin, les informations



Vous êtes le responsable d'un moteur ou d'un annuaire ? Vous travaillez dans le référencement ?
Communiquez et insérez gratuitement votre actualité ici.

19/02/2003 : "IntelliSeek a racheté le métamoteur ProFusion"

La société IntelliSeek Inc., qui propose comme service une technologie de moteur de recherche, a racheté lundi 17 février 2003 le métamoteur ProFusion.com. Le montant de la transaction n'a pas été communiqué. ProFusion a été créé en 1995 à l'Université du Kansas et reçoit aujourd'hui plus de 600 000 visiteurs uniques par mois.

19/02/2003 : "Le moteur de recherche, Altavista, a été vendu à Overture"

Après de graves erreurs stratégiques, Altavista s'est transformé de moteur numéro 1 sur Internet en un parfait inconnu pour des millions d'internautes. Avec la mise à jour effectuée en novembre 2002 (mise à jour esthétique mais aussi technologique), Altavista commençait à reprendre tout doucement une place. Aujourd'hui, nouvelle annonce concernant Altavista. CMGI qui a acheté Altavista en 1998 pour 2,4 milliards de dollars vient de revendre le moteur de recherche 140 millions de dollars (60 millions en cash et 80 millions en action) à Overture. A croire que les fondations de la reprise d'Altavista n'étaient pas aussi solide que ça.

18/02/2003 : "Des weblogs à la mode Google"

« "L'entreprise la plus cool de l'année 2002" (titre décerné par le magazine Fortune) a été rachetée par "la marque de l'année 2002" (titre décerné par le cabinet britannique Interbrand). Pyra Labs, une start-up de San Francisco qui édite Blogger, l'un des plus célèbres outils de création de weblogs, a en effet été acquise par Google, qu'on ne présente plus. Le montant de l'opération n'a pas été divulgué, mais selon Internet.com, celle-ci aurait été réalisée en cash. Sur son blog personnel, Evan Evan Williams, le fondateur de Pyra Labs, estime que cette opération donnera à la start-up "les moyens de construire le projet sur lequel nous travaillons depuis des années". »

15/02/2003 : "KM : Arisem met de la sémantique dans son moteur"

« L'éditeur logiciel français, spécialiste de la gestion des connaissances, ajoute en option un moteur de recherche sémantique à sa solution KM Server. »

15/02/2003 : "Entretien avec Larry Page de Google"

Le journaliste Bradford Delong met à disposition des internautes les notes de son entretien avec Larry Page, le co-fondateur de Google. On découvre que l'objectif de base des cofondateurs n'était pas de créer un moteur de recherche mais une solution pour annoter les sites.

14/02/2003 : "MSN Japon a choisi Overture pour ses liens sponsorisés"

MSN Japon a fait le choix d'Overture pour ses liens sponsorisés.

13/02/2003 : "Les outils de recherche utilisés par les marchands en ligne français"

« Le moteur de recherche fait figure de service stratégique au sein d'un site marchand. A l'instar du rubriquage et autres plans de classement ou taxonomies, il est indispensable aux visiteurs de pouvoir saisir un ou plusieurs mots-clés pour se repérer rapidement dans les méandres des catalogues en ligne... Ce qui n'empêche pas les choix technologiques d'implantation de ce type d'outils d'être forts différents (...). »

13/02/2003 : "Le cloaking... pour ou contre ?"

Un site Web utilise le "cloaking" pour montrer aux moteurs de recherche une information différente que celle vue par les internautes. Est-ce correct ?

13/02/2003 : "Telefonica et Terra Lycos signent pour le long terme"

L'opérateur téléphonique espagnol Telefonica signe un accord sur le long terme avec la société Internet espagnol Terra Lycos. Cet accord va ramener environ 78,5 millions d'euro par an à Terra Lycos et est signé pour une période de 6 ans. Avec cet accord, Telefonica s'engage à faire uniquement la promotion des services de Terra Lycos auprès de ses usagers.

13/02/2003 : "Yahoo! désire se consacrer plus à la recherche"

Les dirigeants de Yahoo! et plus particulièrement son PDG, Terry Semel, ont affirmé que Yahoo! va privilégier la recherche cette année. L'achat d'Inktomi pour 235 millions de dollars donne à Yahoo! l'outil dont il avait besoin.

13/02/2003 : "Yahoo! annonce générer beaucoup de revenus"

Yahoo! Inc., la société propriétaire du célèbre annuaire utilisé par 213 millions de personnes chaque mois, a déclaré que ses revenus risquent de doubler ces 3 à 5 prochaines années et ceci grâce à la publicité et la vente de services aux internautes.

12/02/2003 : "Yahoo et BT ont décidé de joindre leurs forces pour conquérir l'Angleterre"

Yahoo! (société américaine) et BT (British Telecom) ont décidé de proposer des offres communes à leurs clients et aux internautes anglais. Cette offre s'appellera Yahoo! UK Plus. Pour info, Yahoo! a annoncé en début d'année 2003 son désir de devenir fournisseur d'accès en Europe et BT a éjecté, il n'y a pas longtemps, son ancien partenaire... Looksmart.

12/02/2003 : "MSN.fr premier portail du web tricolore ?"

« MSN.fr, le portail média de Microsoft, annonce avoir franchi le cap des 5 millions de visiteurs uniques par mois, dépassant ainsi Yahoo ou Voila. »

12/02/2003 : "Google sacré marque de l'année devant Apple et Coca-Cola"

« Le moteur de recherche sur internet Google a été sacré "marque de l'année", devant Apple et Coca-Cola, selon une étude menée par la société britannique Interbrand. »

11/02/2003 : "Lycos Europe reste dans le rouge"

« Malgré un résultat opérationnel trimestriel positif, le groupe Internet Lycos Europe continue d'afficher de lourdes pertes. »

11/02/2003 : "Societe.com intègre des fonctions d'intelligence économique"

« En partenariat avec Mapstan, le site d'information professionnelle Societe.com intègre désormais des fonctionnalités d'intelligence économique. »

11/02/2003 : "Combien de temps Google va-t-il rester le premier ?"

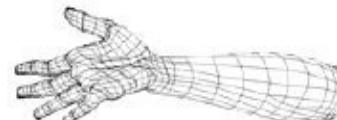
Google ? C'est facile à dire. C'est simple à prononcer. Le site est léger et la recherche rapide. Mais, combien de temps Google va-t-il rester l'enfant chéri des internautes ?

10/02/2003 : "Openfind : Interview de Sun Wu dans MoteurZine 67"

Sun Wu est le fondateur d'Openfind, le moteur de recherche taiwanais qui affiche depuis plusieurs mois 3,5 milliards de pages dans sa base et une version bêta.
« La version finale d'Openfind sera très différente de la version bêta avec une nouvelle interface pour les utilisateurs et de nouveaux services comme la recherche d'images, de sons, d'actualités, de logiciels, etc. J'espère que la version définitive sera en ligne pour juin 2003. »



Les outils de recherche
francophones



19 février 2003 :

[Annuaire des meubles et décoration pour la maison](#)

Annuaire de sites consacrés aux meubles, à l'électroménager et la décoration.

[RPI – Référencement Positionnement Internet](#)

Société de référencement et de positionnement stratégique de sites Internet.

[Echange-liens.com](#)

Annuaire des sites qui souhaitent échanger des liens.

[Cyber-turf](#)

Annuaire des courses hippiques.

[Buffy.fr.to](#)

Annuaire de recherche des sites francophones sur la série télé Buffy.

18 février 2003 :

[Lot-et-Garonne.com](#)

Annuaire des villes et villages du Lot et Garonne en Aquitaine.

[Annuaire des Côtes d'Armor](#)

Les sites web en côtes d'armor, sites internet costarmoricaïns.

15 février 2003 :

[Atoutsref](#)

Société de référencement basé à Perpignan.

14 février 2003 :

[Qualitrafic](#)

Solutions globales de référencement et génération de trafic.

[Lemanique.com](#)

le portail de la région franco-italo-suisse du commerce.

[Annuaire des gîtes et des chambres d'hôtes](#)

Annuaire et répertoire de recherche de locations vacances de types gîtes d'étape, chambres d'hôtes, chalets et hébergements divers à travers la France.

12 février 2003 :

[Topclito](#)

Annuaire x avec de nombreuses catégories, les sites certifiés par Topclito sont de qualité.

11 février 2003 :

[Le médocain](#)

Annuaire des sites sur le Médoc et le vin.

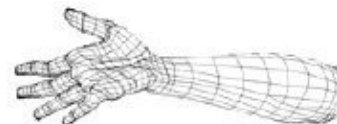
10 février 2003 :

[Lekti-écriture.com](#)

Annuaire des sites d'art et d'artistes sur Internet.



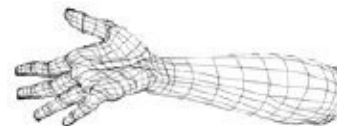
Espace publicitaire



– vosre publicité lue par 15.000 personnes ! –



Conclusion



ATTENTION

La liste des abonnés n'est pas disponible. Nous ne la donnons ni ne la revendons à personne. En revanche, vous pouvez sponsoriser notre lettre d'information : contactez-nous pour discuter des modalités.

[Abonnement gratuit](#)
à « Moteurzine »

[Désinscription de](#)
« Moteurzine »

les crédits

Chronique sur « Google »

[Olivier DUFFEZ](#) de
[WebRankInfo](#)

Chronique sur les moteurs

[Gilbert WAYENBORGH](#)
de [DeepIndex](#)

Chronique sur le Web
invisible

[Christophe ASSELIN](#) de
[Intelligence-center.com](#)

Infodesign

[Max](#), le talentueux !

Le reste (mise en page,
édito, actualités, les
nouveauités et l'interview)

[CHRIS HÉDÉ](#) de
[MoteurZine / Enfin](#)