

## Plagiat et référencement : comment détecter les problèmes ? Quels sont les recours ? (2ème partie : les flux RSS)

[Retour au sommaire de la lettre](#)

Plusieurs éditeurs de sites web se sont aperçus, dans le passé, que des webmasters inconvenants copiaient purement et simplement leur contenu en pensant bénéficier d'un bon référencement obtenu par le site "source" parfois après de longs mois, voire années, de travail. Ce type de copie illicite s'effectuant bien entendu sans demander la moindre autorisation. Comment détecter ces plagiat ? Comment en découvrir les auteurs ? Comment faire pour stopper ces agissements ? Cet article tente d'apporter quelques éclairages à ces questions cruciales avec ce mois-ci un focus sur la question des flux RSS qui peuvent avoir un effet négatif sur un référencement...

### Préambule

Dans l'article du mois dernier, nous avons abordé les problèmes liés au plagiat d'une page web et quels étaient les moyens de les détecter. La reprise automatique de contenu (ou syndication de contenu) d'un site à un autre s'est également développée grâce à l'essor des blogs et l'utilisation de plus en plus importante des flux (ou fils ou canal) RSS. Comment cela fonctionne-t-il ? Quels sont les avantages liés à l'utilisation des flux RSS ? Quels sont les inconvénients ? Existe-t-il une vraie limite entre plagiat et contenu dupliqué ? C'est ce que nous allons développer au cours de cet article.

### Comment fonctionne un flux RSS ?

Les flux RSS (*Real Simple Syndication* entre autre signification de l'acronyme) permettent de publier un contenu (ou une partie de ce contenu) sur d'autres sites de manière automatique et instantanée. Ces fichiers XML sont mis automatiquement à jour en temps réel.


Simple fichier texte stocké sur le serveur d'un site Internet, disposant de sa propre URL, un flux RSS contient généralement un titre (title), la date de publication, une partie de la description (description) de l'article ainsi que l'url de l'article désigné (link).

```
<?xml version="1.0" encoding="iso-8859-1" ?>
<rss version="2.0" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
- <channel>
  <title>Le Monde.fr : A la une</title>
  <link>http://www.lemonde.fr</link>
  <description>Toute l'actualité au moment de la connexion</description>
  <copyright>Copyright Le Monde.fr</copyright>
- <image>
  <url>http://medias.lemonde.fr/mmpub/img/lgo/lemondefr_rss.gif</url>
  <title>Le Monde.fr</title>
  <link>http://www.lemonde.fr</link>
</image>
  <pubDate>Tue, 10 Apr 2007 16:38:02 GMT</pubDate>
- <item>
  <title>Bagdad, quatre ans après : la partition confessionnelle</title>
  <link>http://www.lemonde.fr/web/article/0,1-0@2-3218,36-893996,0.html?xtor=RSS-3208</link>
  <description>Sous l'effet des attentats et des affrontements entre musulmans sunnites et chiites, la division de la capitale irakienne en deux blocs est pratiquement achevée. Seule une partie du centre-ville conserve encore une certaine mixité.</description>
  <pubDate>Tue, 10 Apr 2007 15:30:45 GMT</pubDate>
  <guid isPermaLink="false">http://www.lemonde.fr/web/article/0,1-0@2-3218,36-893996,0.html?xtor=RSS-3208</guid>
  <enclosure url="http://medias.lemonde.fr/mmpub/img/lgo/lemondefr_rss.gif" type="image/gif" length="9150" />
</item>
</channel>
</rss>
```

Pour consulter un fils RSS, il faut utiliser un outil spécifique, appelé "lecteur RSS" ou "agrégateur". Celui-ci peut être intégré à un navigateur Internet (**Mozilla Firefox**), soit autonome (**Feedreader**) ou encore être intégré à un logiciel de messagerie (**Mozilla Thunderbird**).

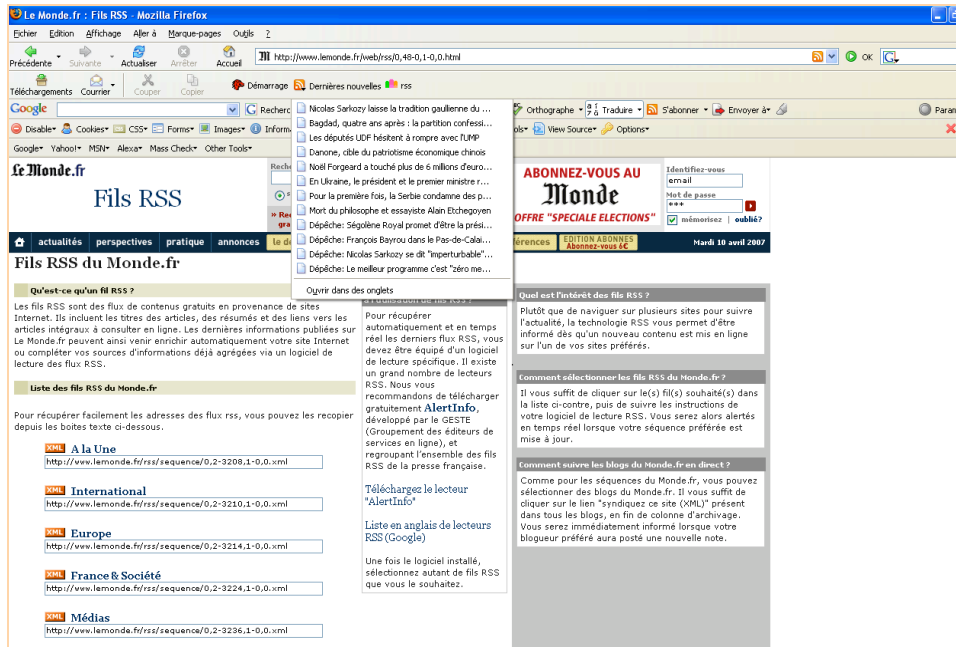
Vous trouverez une liste de lecteurs RSS non exhaustive sur dmoz.org :

[http://directory.google.com/Top/Reference/Libraries/Library\\_and\\_Information\\_Science/Technical\\_Services/Cataloguing/Metadata/RDF/Applications/RSS/News\\_Readers/](http://directory.google.com/Top/Reference/Libraries/Library_and_Information_Science/Technical_Services/Cataloguing/Metadata/RDF/Applications/RSS/News_Readers/)

Un navigateur Web tel que **Mozilla Firefox** est en mesure de repérer l'adresse des flux RSS. L'internaute clique juste sur l'icône  qui apparaît au sein de la barre de navigation pour s'abonner au flux.



Le navigateur crée alors un marque-page dynamique sur lequel il suffira ensuite de cliquer pour afficher les titres correspondant.



Le menu déroulant permettra ensuite d'accéder directement à l'article souhaité.

Le Monde.fr : Nicolas Sarkozy laisse la tradition gaullienne du référendum à ses rivaux - Mozilla Firefox

http://www.lemonde.fr/web/article/0,1-002-823446,36-893964,0.html?xtor=RSS-3208

Nicolas Sarkozy laisse la tradition gaullienne du référendum à ses rivaux

LE MONDE | 10.04.07 | 16h50 • Mis à jour le 10.04.07 | 16h50

Suivez l'information en continu, accédez à 70 fils de dépêches thématiques.  
Abonnez-vous au Monde.fr : 6€ par mois + 30 jours offerts

C'est l'un des multiples brouillages de cette campagne présidentielle. Le lointain héritier du gaullisme, qui avait fait largement usage du référendum dans une pratique plébiscitaire, promet de ne pas y avoir fréquemment recours: le candidat de l'UMP, Nicolas Sarkozy, a ainsi indiqué qu'il fera, s'il est élu président, un usage "modeste" de ce type de consultation populaire. En revanche, ses deux principaux rivaux - qui "descendent" de familles politiques de tradition plus parlementariste - ont annoncé l'un et l'autre qu'en cas de victoire ils consulteraient les Français sur plusieurs sujets majeurs.

M. Sarkozy a justifié ses réserves à ce sujet dans un entretien à l'hebdomadaire L'Express du 27 mars. "Le quinquennat a beaucoup changé les choses", a expliqué le candidat de l'UMP. Avant de s'interroger en ces termes: "Il y a la présidentielle, puis les législatives, se profilent ensuite les municipales, les cantonales et les régionales. Croyez-vous que, si je suis élu, je vais aussitôt dire aux Français: "Excusez-moi, j'ai besoin de vous demander votre avis sur un autre sujet"?"

François Bayrou (UDF) se dit au contraire "partisan de réhabiliter le référendum pour donner la base populaire nécessaire à de grands projets". Pour le candidat oestriste, "le référendum, sur un grand sujet, c'est le moyen de donner conscience des enjeux et d'entraîner la mobilisation. Le référendum place le citoyen en position de responsabilité, de gouvernant. Et non pas en position de consommateur qui ne pense qu'à sa situation personnelle."

Nicolas Sarkozy a indiqué qu'il fera, s'il est élu président, un usage "modeste" du référendum.

Réagir à cet article Voir

Nicolas Sarkozy laisse la tradition gaullienne du référendum à ses rivaux

Votre réaction

### Quels sont les avantages liés à l'utilisation des flux RSS pour un référencement ?

La technologie RSS possède des avantages non négligeables par rapport à d'autres canaux de communication :

#### - L'information est transmise directement à l'utilisateur

Un internaute qui reçoit lui-même l'information qu'il a demandée sera beaucoup plus réceptif sur le message/l'information reçu. L'utilisation du flux RSS passe obligatoirement par une demande d'abonnement.

#### - La reprise de contenu déjà existant via des flux RSS peut être un bon moyen de faire vivre son site et/ou d'éviter la Sandbox

Les moteurs de recherche apprécient l'ancienneté. Les nouveaux sites doivent faire leur preuve et peuvent être mis en quarantaine (phénomène de Sandbox) en attendant d'avoir suffisamment d'historique, de liens externes pour remonter dans les résultats naturels. La reprise de flux RSS en provenance de sites qui font référence dans leur domaine d'activité peut permettre au diffuseur d'obtenir un historique en attendant d'avoir son propre contenu. De plus, cela peut faire vivre son site et indiquer aux robots des moteurs que les pages sont souvent mises à jour...

#### - Les flux RSS peuvent apporter de nombreux backlinks (liens externes).

Pour juger de la performance d'un flux RSS, au niveau du référencement, il suffit d'essayer et d'effectuer une recherche sur un moteur de recherche. Prenons l'exemple du basket et d'une requête du type « Detroit domine Cleveland » sur Google pour mesurer l'impact d'un fil RSS.

Web Résultats 1 - 10 sur un total d'environ 60 400 pour **Detroit domine Cleveland** (0,12 secondes)

1 résultat stockés sur votre ordinateur - Masquer - À propos  
contenu-duplique-Avri200... - NBA **Detroit domine Cleveland** - 17:29

Football avec Le Figaro sports : résultat football, calendrier ...  
**Detroit domine Cleveland**. A domicile, les Pistons ont remporté le choc de la Conférence Est qui les opposait à **Cleveland** (87-82). **Detroit** est bien le patron ...  
sports.lefigaro.fr/infobasket\_ball\_nba\_detroit\_domine\_cleveland\_55709.html - 8 avr 2007 - Pages similaires

Football avec Le Figaro sports : résultat football, calendrier ...  
Si certains souhaitent encore titiller **Detroit** pour le leadership de la Conférence ... **Detroit domine Cleveland**. [08/04 16h41] Basket-ball, Pro A, 29e j ...  
sports.lefigaro.fr/article\_basket\_ball\_a\_l\_est\_du\_nouveau\_15367.html - 8 avr 2007 - Pages similaires

**Basket-ball, NBA : Detroit domine Miami | Actuline**  
Basket-ball, NBA [07/04/2006 08h03] **Detroit domine Miami** : Dans le choc de la ... NBA : Denver **domine Cleveland** - Basket-ball, NBA : Miami **domine Detroit** ...  
www.actuline.com/basket\_ball\_nba\_detroit\_domine\_miami.html\_0 - 15k -  
En cache - Pages similaires

Si certains souhaitent encore titiller **Detroit** pour le leadership de la ... 21h29 Basket-ball, NBA : **Detroit domine Cleveland**, 16h41 Basket-ball, Pro A, ...  
www.sport24.com/  
sport24\_article\_actualite\_sportive\_basket\_ball\_nba\_compte\_rendu\_de\_la\_nuit\_a\_l\_est\_du\_n...  
- 8 avr 2007 - Pages similaires

Le quatrième site (actuline) est un agrégateur de flux RSS. Il a pour vocation de reprendre le contenu de sites partenaires via des flux RSS.

The screenshot shows the Actuline website interface. At the top, there is a search bar with the text "Rechercher : " and a "Recherche" button. Below the search bar is a navigation menu with "Populaire", "Tags", and "Top Vote". The main content area features a news article titled "Basket-ball, NBA : Detroit domine Miami" with a sub-header "Le choc de la Conférence Est a tourné à l'avantage des Pistons de Detroit, vainqueurs du Heat de Miami (94-88)". The article includes a date "Le 01/04/2007 à 23:39:57 - Catégorie : Sport" and tags "basket ball nba | pistons de detroit | basket nba | miami | domine | avantage | choc |". Below the article, there are several related links from other sites, such as "Football avec Le Figaro sports : résultat football, calendrier ..." and "Basket: Miami domine Detroit en NBA - Yahoo! Sport ...". On the right side, there is a "Categories" list including "Actualité", "Automobile", "Culture", "Economie", "High Tech", "Insolite", "Jeux", "Media", "People", "Politique", "Science", and "Sport". Below the categories is a "Tags Actuline" section with various tags like "premier ministre euros", "lemonde fr", "la campagne finale jeu", "ségotène royal homme", "rencontre rugby vi nations", "paris face elle", and "mondial rugby encore".

Cette page présente l'avantage de proposer des liens en dur vers des sites en étroite relation avec le contenu de la page. Les liens figurant sur cette page sont donc considérés comme très pertinents par le moteur de recherche.

### Quels sont les inconvénients liés à l'utilisation d'un flux RSS ?

L'éditeur d'un flux RSS doit faire face à un certain nombre de contraintes avant d'en tirer un plein bénéfice :

- Chaque diffuseur qui décide d'insérer le flux RSS apportera des liens directs au site éditeur sous condition que ce flux puisse être affiché avec un script parseur XML générant des liens en dur (HTML simple).

Il est très important que les liens (internes ou externes) soient en dur et non cryptés en format Javascript sinon ils ne seront pas suivis par les moteurs de recherche.

- Pour obtenir la diffusion du flux auprès d'un maximum de sites, il est indispensable de le faire connaître et d'engager une campagne d'inscription au sein d'annuaires spécialisés.

Le référencement passe par un principe de base qu'il convient de garder en tête : faire connaître un site Internet auprès de ses pairs lors du lancement du site mais également lors du lancement de

nouvelles fonctionnalités. Si un site lance un flux RSS, il faut impérativement inscrire le flux sur des annuaires pour qu'un maximum de diffuseurs puissent le connaître et l'utiliser sur leur propre site. Pour ce faire, il existe des annuaires comme **Eco Netvibes** (<http://eco.netvibes.com/submit/>) ou **La mooche** (<http://www.lamooche.com/>).

**- Il est prudent d'éviter de proposer l'article dans son intégralité auprès des partenaires afin d'éviter toute duplication de contenu.**

Un contenu ne doit pas être présent sur plusieurs URLs différentes. Ainsi, un moteur de recherche n'appréciera pas une page d'accueil disponible sur plusieurs adresses (exemple : [www.nom-de-domaine.com/index.php](http://www.nom-de-domaine.com/index.php) et [www.nom-de-domaine.com](http://www.nom-de-domaine.com)) et risquera de déclasser le site de ses propres résultats. Il est indispensable de ne pas avoir deux pages avec un contenu strictement identique. Lorsqu'il s'agit d'un flux RSS et que le diffuseur a pris soin de ne reprendre qu'une partie des données (par exemple 10% de l'article de l'éditeur), le moteur arrivera à faire la différence.

**- Si un flux RSS est distribué sur d'autres sites, il faut s'assurer qu'ils incluent un lien direct vers l'article original.**

Il importe que la source d'origine soit clairement affichée. Cela évitera ainsi une remontrance éventuelle de la part de l'éditeur tout en jouant sur l'indice de popularité de la page.

Comme on a pu le voir, l'utilisation d'un flux RSS est tout à fait bénéfique pour le référencement et les contraintes restent tout à fait maîtrisables.

Les flux RSS permettent de récupérer et dupliquer l'information automatiquement. Mais existe-t-il des règles juridiques précises concernant la reprise de contenu ? Comment faire pour se protéger ?

**Quelles sont les règles juridiques appliquées à la reprise de contenu ?**

Il appartient à la personne reprenant du contenu en provenant d'un tiers (via un fil RSS par exemple) sur son propre site internet de bien vérifier sous quelles conditions chaque flux est publié et si sa copie est autorisée (pour un usage commercial ou non).

L'éditeur peut avoir fait appel à une licence « open source » (comme *Creative Commons*). La licence propose gratuitement des contrats flexibles de droit d'auteur pour diffuser des créations. En l'absence de licence, c'est le copyright qui s'applique.

Il est relativement difficile de définir une limite entre reprise de contenu et plagiat. Sans notification explicite, il est fortement recommandé de prendre contact avec les auteurs d'un texte avant de l'utiliser. Il est impératif d'épuiser tous les recours à l'amiable avant d'entamer un procès (qui peut être onéreux et complexe surtout si le site incriminé est hébergé à l'étranger).

En tout état de cause, si un site web met en ligne un flux RSS, cela peut signifier qu'il va servir aux internautes détenteurs d'un agrégateur de flux, mais vous n'avez pas pour autant la permission de reprendre ces données sur votre site web... Vérifiez bien que cette information est fournie en clair sur le site qui propose le flux... Entre usage privé et mise en ligne publique, il y a une marge que vous n'êtes pas obligatoirement autorisé à franchir...

Un premier contact (par mail par exemple) constatant l'infraction peut débloquent de nombreuses situations. Sans réponse, une lettre recommandée pourra être envoyée. L'adresse du plagieur pourra être obtenue en effectuant un whois sur le nom de domaine (cf article du mois précédent). Suite à cela, sans réponse, on pourra faire appel à un huissier de justice et réclamer des indemnités compensatoires. Toujours sans réponse, le plaignant pourra porter plainte pour contrefaçon et se constituer partie civile pour demander réparation financière.

Pour éviter tout problème, il vaut mieux, dans la mesure du possible, s'assurer de l'antériorité d'un document. Pour ce faire, il existe plusieurs possibilités. Le texte peut être déposé auprès d'une société d'auteur comme **La Société des Compositeurs et Auteurs Multimédias**. Le texte peut être déposé auprès d'un représentant légal de l'état (huissier de justice ou notaire). Cependant, les coûts les moins onéreux restent l'envoi à soi-même d'un courrier cacheté en recommandé, ou l'envoi par la poste à l'INPI (Institut national de la propriété industrielle) d'une enveloppe **Soleau**.

Dans toutes les situations, le dialogue et l'échange peuvent souvent résoudre des situations bien compliquées comme le montre le conflit qui a opposé les moteurs de recherche (Google et Msn) à l'association Copiepresse regroupant de nombreux journaux belges.

### ***Quels sont les exemples de conflit rencontrés par les diffuseurs de flux RSS ?***

« Google news » a pour vocation de reprendre des milliers d'articles (même si, dans ce cas, la récupération ne s'effectue pas techniquement via un flux RSS, l'exemple n'en reste pas moins intéressant et pertinent) en provenance de la presse du monde entier. Cela peut se faire sans en demander la permission mais, quoiqu'il arrive, il n'y aura pas de rémunération.

Cela n'a pas été du goût des éditeurs de plusieurs journaux belges dont Le soir (<http://www.lesoir.be/>) et La Libre Belgique (<http://www.lalibre.be/>). En septembre 2006, ces journaux, associés à Copiepresse qui gère les droits d'auteurs des journalistes belges, ont porté plainte contre Google. Ils refusaient de voir leur contenu indexé par le célèbre moteur de recherche dans sa partie « News » sans contrepartie ni autorisation.

Le tribunal a accepté la demande de la presse belge et a demandé expressément à Google de retirer toute référence (contenu, image, graphisme...) à ces journaux dans sa partie News sous peine d'amende (de l'ordre de 500 000 d'euros par jour de retard !)

Google a accepté la sentence. Étrangement, quelques semaines plus tard, les sites de ces journaux belges se sont retrouvés complètement désindexés des résultats naturels du moteur de recherche américain. Une recherche « site :www.lesoir.be » sur le moteur ne laisse apparaître aucune page indexée alors que le site existe bien réellement.

L'histoire faillit se répéter avec Msn. En octobre 2006, le moteur fut accusé de récupérer le contenu de sites belges sans contrepartie. Mais, face aux enjeux, le moteur et les éditeurs belges se mirent vite d'accord. À croire que les droits des flux RSS, c'est une histoire belge ! En tout état de cause, sachez que le fait que des informations soient à votre disposition sous la forme de flux de données ne vous permet pas de faire n'importe quoi...

**Damien Henckès, Consultant, Nextedia**