

10 outils de surveillance de sites web

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Dans le cadre d'un processus de veille, on a souvent besoin d'être averti lorsqu'une page web, se trouvant sur un site distant, est modifiée. Il existe ainsi de nombreux outils permettant de générer un flux RSS ou d'envoyer un e-mail à chaque modification ainsi constatée. Certains de ces outils sont très performants et permettent, de façon gratuite ou peu onéreuse, de mettre en place des procédures de surveillance très pointues. Chacun a cependant ses avantages et ses inconvénients... Voici 10 exemples de sites web qui ont pour ambition de faciliter vos travaux de veille, décrits avec précision pour vous permettre de faire votre choix...

Le principe des outils que nous allons présenter dans cet article est de vous avertir dès qu'un changement est intervenu sur un document visible sur le Web. D'une part, ils vous permettent de générer une alerte e-mail ou de créer un flux RSS à partir de pages web ou de sites qui n'en proposent pas. D'autre part, ils permettent de resserrer les alertes à une portion spécifique d'une page web. Ils fonctionnent donc à mi-chemin entre agents de veille et applications de *Web scrapping*. Rappelons que ce terme désigne toute sorte de moyens permettant d'extraire un contenu web afin de l'utiliser à d'autres fins. On peut distinguer différentes fonctions qui vont nous permettre de les différencier :

- L'existence ou non d'une fonction d'alerte ;
- Le type d'alerte générée : contenu web, e-mail, flux RSS ;
- La fréquence des alertes ;
- La taille maximale des fichiers qu'il est possible de surveiller ;
- La possibilité de simuler des requêtes en mode Post ou en mode Get et donc de surveiller un formulaire ;
- La possibilité d'archiver les différentes versions des pages à surveiller ;
- La possibilité de filtrer le contenu de la page cible ;
- La possibilité de tracer des modifications intervenus sur des documents non HTML ;
- La possibilité de faire de la surveillance de pages en FTP, nécessitant l'utilisation de protocoles SSL ou HTTPS ;
- La possibilité d'utiliser des adresses URL contenant des variables ;
- La possibilité de modifier le type d'agent utilisateur qui sera envoyé ;
- L'utilisation d'un mode Proxy.

Les applications que nous allons examiner ici présentent toutes l'avantage d'être gratuites ou proposées à un coût très réduit.

Nos tests porteront sur des adresses URL générées quand on lance une recherche avancée dans Google en filtrant les résultats sur les dernières 24 heures (ou dernière semaine, etc.) En bref, le principe est de générer un flux RSS à partir d'une recherche dans Google limitée dans le temps. Nous verrons ici et là quelques cas particuliers d'utilisation...

Change Detect

Accédez à cette page web pour utiliser l'outil **Change Detect** : <http://www.changedetect.com>. Une fois que vous avez procédé à votre inscription, il vous suffira ensuite de cliquer sur le lien *Login*.

Cette application peut être testée durant une période de 30 jours. Les fonctionnalités qui sont offertes sont toutes visibles sur cette page :

<http://www.changedetect.com/changedetect/features.asp>.

Une comparaison entre les différentes versions est proposée à cette adresse :

<http://www.changedetect.com/changedetect/plan-comparison.asp> (de 1.95\$ à 39.95\$ par mois).

En version d'essai, peu d'options sont disponibles mais voici un exemple d'utilisation :

1. Cliquez sur *Service Management - Create New Monitor*.
2. Saisissez l'adresse URL de la page puis cliquez sur *Submit*.

Par exemple, nous allons surveiller tous les *curriculum-vitae* qui ont été postés sur le site DoYouBuzz, qui datent de moins d'une semaine et qui concernent des chefs de projet web basés sur Paris. Etant donné qu'il n'existe pas de module de recherche et pas, non plus, de flux RSS sur ce site, nous allons utiliser cette requête dans Google : [paris "chef de projet" web site:http://www.doyoubuzz.com](http://www.doyoubuzz.com). La seule astuce consiste à utiliser les paramètres de recherche avancée et de restreindre les résultats à la semaine dernière. Au final, on obtient ce type d'adresse URL :

[http://www.google.fr/search?hl=fr&q=paris+ "chef+de+projet+web"+site:http://www.doyoubuzz.com&as_qdr=w](http://www.google.fr/search?hl=fr&q=paris+).

Vous allez voir ce type d'indication sur Change Detect : #211900 (*needs a name*).

3. Cliquez sur ce lien puis sur le bouton *Modify*.
4. Définissez un titre et une catégorie.
5. Saisissez éventuellement une description ainsi que l'adresse URL de référence.
6. Cochez la case *Stealth*.

Les agents utilisateurs (dont le robot de Change Detect) envoient automatiquement un tag d'identification au serveur qu'il visite. En mode "Stealth", aucune identification n'est assurée.

7. Cochez la case *Silent* si vous ne souhaitez pas recevoir d'alerte par e-mail.
8. Cochez la case *Send cd-diff file* : quand cette option est cochée, Change Detect enverra un exemplaire de la page qui a été modifiée avec, en surbrillance, les éléments de la page qui ont changés.

La liste déroulante *Error Handling* vous offre le choix entre trois options :

- *Default* : la moindre source d'erreur vous sera notifiée par e-mail ;
- *Ignore Errors* : aucune erreur ne sera indiquée ;
- *Errors Only* : seules les erreurs vous seront notifiées.

La rubrique *Notification Triggers* permet de poser des filtres supplémentaires :

- *If the page size changes by at least "x" bytes* : permet de définir à partir de combien d'octets l'application considèrera qu'un changement est intervenu sur la page.
- *When all of the words are detected* : quand tous les mots suivants auront été détectés.
- *When the exact phrase is detected* : quand cette expression exacte aura été détectée ;
- *When at least one of the words is detected* : quand au moins un de ces mots aura été détecté ;
- *When one of the words is not found* : quand un de ces mots n'est plus trouvé.

9. Cliquez sur le bouton *Submit*.

Records 1 to 6 of 6

Category Filter: -Show All-

[New Category] | Move | Bulk Management | Default | Delete

Bulk actions are not supported for your current membership plan. [Order Now](#).

<input type="checkbox"/>	#GVMDTCWeb Page	Last Changed	F	S	Q	±	E
<input type="checkbox"/>	1 ChangeDetect	February 18, 2009	24				
<input type="checkbox"/>	2 Directory Mozilla	June 24, 2008	24				
<input type="checkbox"/>	3 Google Chrome	January 16, 2009	12				
<input type="checkbox"/>	4 Google Directory	August 02, 2006	168				
<input type="checkbox"/>	5 Nonsensical domain name that errors	October 30, 2007	720				
<input type="checkbox"/>	6 Spy On Your Competitors	March 13, 2009	24				

Vous allez retrouver votre tableau de bord avec les indications suivantes :

- *G* : accès à la page source ;
- *V* : accès à la page des paramètres ;
- *M* : modification de la page des paramètres ;
- *D* : suppression de la page à surveiller ;

- T : permet de tester la validité d'une URL ;
- C : lance une vérification manuelle de la page source ;
- F : affiche la fréquence d'actualisation ;
- S : indique si le mode "Stealth" est activé ;
- Q : indique si le mode silencieux est activé ;
- ± : indique si le mode "cd-diff" est activé ;
- E : affiche d'éventuelles erreurs survenues durant le processus de monitoring.

Afin d'ajouter une liste d'adresses URL, cliquez sur *Service Management - Add Bulk monitors*. Dans l'interface qui apparaît, vous pouvez copier le contenu d'une page qui contient une liste de liens. La limite est de 5000 caractères. En dépit d'une interface un peu alambiquée, ChangeDetect s'avère être un excellent outil de monitoring...

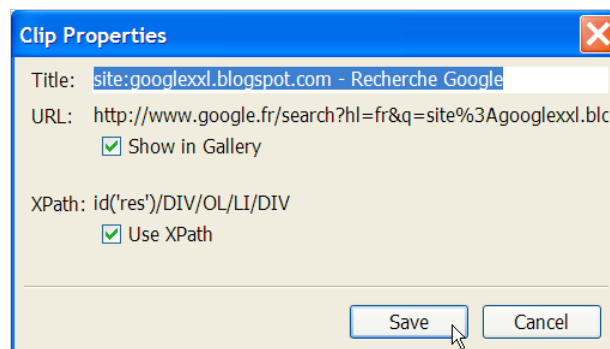
Fireclip

Fireclip est une extension Firefox (2.0 - 3.2a1pre) qui vous permet de traquer les changements survenue sur une portion d'une page web.

1. Afin de procéder à l'installation de cette extension, accédez à cette adresse : <https://addons.mozilla.org/en-US/firefox/addon/5850>.
2. Affichez ensuite la page web à surveiller.
3. Servez-vous de la combinaison de touches *Alt + Q* afin d'afficher l'interface de Fireclip dans le volet latéral.
4. Cliquez sur le bouton *Select Clip*.
5. Sélectionnez la portion de la page à surveiller puis double-cliquez sur la sélection afin de l'ajouter à votre galerie.



6. Saisissez un nom.
7. En vous servant du menu contextuel, cliquez sur le sous-menu *Properties*. La requête XPath qui a été utilisée sera affichée.



En admettant que nous souhaitions savoir toutes les nouvelles pages publiées par un site, nous utiliserons dans Google cette requête : `site:exemple.fr` et limiterons les recherches aux dernières 24 heures. La requête XPath utilisée est la suivante : `id("res")/DIV/OL/LI/DIV`.

Deux mots d'explications sont ici nécessaires : une expression XPath est composée de sélecteurs qui définissent chacun un ensemble de nœuds. Il existe un cours très intéressant sur le site de Jean-Luc Massat : <http://www.dil.univ-mrs.fr/~massat/ens/xml/4-xpath.html>. Si, dans un premier temps, le langage XPath permet d'identifier un ensemble de nœuds dans un document XML, il vous servira aussi à extraire un contenu sélectionné à l'intérieur d'une page web. Par exemple et afin d'afficher tous les liens d'une page web, il vous suffira d'utiliser cette expression : `//a/@href`. Procédez à un test rapide en utilisant Google Tableur et en posant cette formule : `=importXml(www.google.fr;"//a/@href")`.

Vous pouvez faire le test en utilisant une autre extension pour Firefox qui permet de trouver automatiquement la requête XPath quelque soit l'élément de page web considérée : **Xpath Checker** : <https://addons.mozilla.org/en-US/firefox/addon/1095>. Notez qu'une excellente alternative est proposée par cette autre extension Firefox : **Update Scanner** (<https://addons.mozilla.org/fr/firefox/addon/3362>).

Feedity

Feedity (<http://feedity.com/builder.aspx>) permet de générer un flux RSS à partir d'une page web. Procédez à votre inscription sur ce service avant de commencer à l'utiliser. Le principe est très simple :

1. Saisissez l'adresse URL générée par votre recherche puis cliquez sur le bouton *Preview*.
2. Cliquez sur le bouton *Advanced Refine* afin de parfaire le contenu de votre flux RSS.

En admettant que l'URL soit celle-ci :

http://www.google.fr/search?as_epq=jean+noel+anderruthy&as_qdr=d.

Quand on affiche le code source de la page, les liens sont tous annoncés par un début de balise `<h3`. Il nous suffit donc de saisir ces deux paramètres dans les zones de texte *Start Block* et *End Block* : `<h3` et `</h3>`. Le principe est donc différent d'une requête XPath puisque l'on n'utilise pas de sélecteurs.

The screenshot shows the Feedity interface for a Google search. On the left, under 'Preview Results', there is a list of search results including 'Amazon.fr: Depuis 3 mois - Internet / Sciences, Techniques et ...', 'Amazon.fr: Depuis 1 mois - Sciences, Techniques et Médecine / De ...', 'Livres Informatique - Librairie Eyrolles', 'Livres Audionumérique - Librairie Eyrolles', 'Google XXL: Personnaliser un formulaire Google', and 'Docs'. Below this list, there is a button 'Update Feed >>' and a dropdown menu set to 'News & Culture'. On the right, the 'Advanced Refine' panel is active, showing the configuration for the RSS feed. It includes a 'Refine' button and a 'Retrieve Published Dates' section.

Par ailleurs, l'inconvénient de cette méthode est que l'élément qui sera capturé sera visible sous la forme d'un lien unique. Dans la version gratuite, on ne peut pas créer un flux RSS "avancé" avec un titre, un lien et un contenu qui soit clairement séparé. Un comparatif des différentes versions est visible à cette adresse : <http://feedity.com/plans.aspx> (de 39 \$ à 369 \$ par an). Ceci dit, Feedity est un service qui fonctionne parfaitement...

Feed43

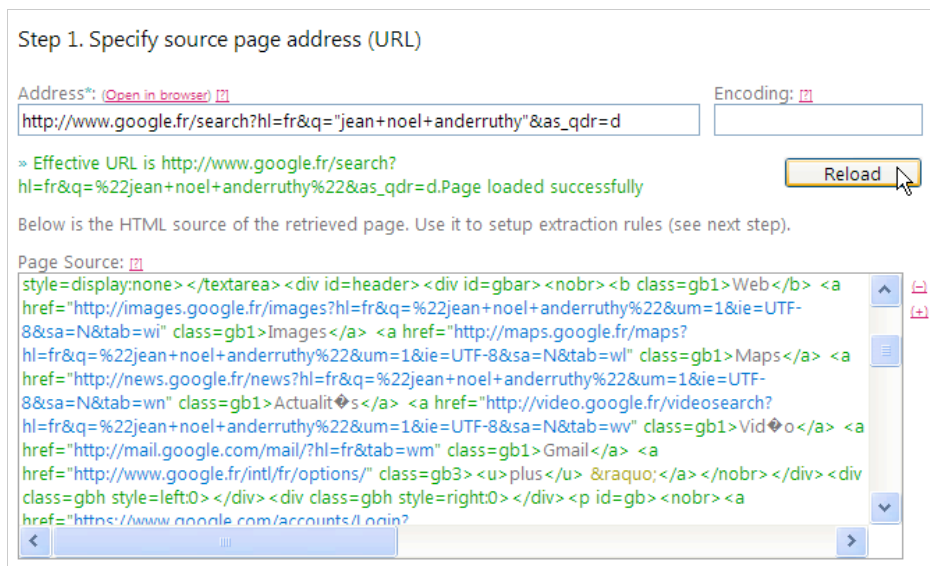
Feed43 (*Feed For Free*) est accessible à partir de cette adresse : <http://feed43.com>. Un comparatif entre la version gratuite et les versions payantes (de 29\$ à 89\$ par an) est visible sur cette page : <http://feed43.com/upgrade.html>. Concernant la version gratuite les conditions sont les suivantes :

- Le nombre de flux qu'il vous sera possible de générer est illimité ;
- Une inscription "*Delivered by Feed43 service*" sera ajoutée à chacun des éléments qui composent votre flux ;
- L'intervalle de mise à jour est de 6 heures ;
- La page source ne doit pas excéder une taille de 100 Mo ;
- Le nombre maximal d'éléments qui composent votre flux sera de 20 ;
- Le nombre maximal de macros qu'il vous sera possible d'appliquer à votre flux sera de 20.

1. Cliquez sur le lien *Create new feed*.
2. Cochez la case *I agree to the Terms of Service* puis cliquez sur *Continue*.
3. Cliquez sur le lien *Switch to advanced mode* afin d'afficher l'ensemble des options qui sont à votre disposition :

- *Address* : adresse URL de la page web ;
- *Encoding* : par défaut, Feed43 utilise l'encodage UTF-8 (Unicode) pour les flux RSS qui seront générés. Si la page web qui sert de source n'est pas encodée de cette manière, Feed43 essaiera automatiquement de l'encoder en UTF-8. Si, pour différentes raisons, le processus d'encodage échoue, vous avez la possibilité de spécifier manuellement la page de codes à utiliser. La syntaxe à utiliser sera de ce type : `iso-8859-1%>`.

Une fois que vous avez indiqué la page web voulue, cliquez sur le bouton *Reload* afin d'afficher le code source de la page.



L'intérêt de cette interface est que les balises importantes sont clairement mises en surbrillance.

La rubrique *Define extraction rules* permet de définir des règles d'extraction.

Ces macros sont autorisées lors de la création du schéma d'extraction à appliquer à la page web source :

- `{%}` : correspond à n'importe quelle séquence de texte et l'ajoute aux résultats ;
- `{*}` : correspond à n'importe quelle séquence de texte et l'exclue des résultats ;
- `{_}` : correspond à n'importe quelle séquence d'espaces (dont les sauts de ligne) et l'exclue des résultats ;
- `{|}` : correspond à n'importe quel saut de ligne et l'exclue des résultats.

Deux zones de texte sont visibles :

- *Global Search Pattern* : permet de définir la portion de la page web qui vous intéresse ;

- *Item (repeatable) Search Pattern* : permet de définir les éléments qui se répètent et qui vont constituer le contenu de votre flux RSS.

Nous allons de nouveau utiliser la page des résultats de Google et utiliser ce schéma d'extraction : `<h3 class=r>{&}</h3>{*}<div class={&}`

En bref, on récupère tout de qui est contenu entre les balises `<h3>` puis entre la balise `<div class>` et ``.

Vous allez avoir cette mention : *OK (n items found)* ainsi qu'un aperçu des résultats dans la zone de texte *Clipped Data*.

Step 2. Define extraction rules

Global Search Pattern:

Item (repeatable) Search Pattern*:

» OK (9 items found)

Below is list of extracted text snippets ({{N}}). You can reference them when setting up item properties (see next step).

Clipped Data:

Item 2
{%1} =
{%2} = "s hc">Du Web 2.0 au Web 3.0 - Les nouveaux services Internet. par Jean-Noël Anderruthy

Item 3
{%1} =
{%2} = "s">Google AdWords, Analytics et AdSense - Gagnez plus de trafic sur votre site; De Je

Item 4
{%1} =
{%2} = "s hc">Blogs, podcasts, RSS; De Jean-Noël Anderruthy

Puisque nous avons scindé le processus d'extraction en deux parties, nous avons ces deux indications : `{%1}` et `{%2}`.

Il nous faut maintenant définir un format de sortie en utilisant les zones de texte visibles sous la rubrique *RSS item properties* :

- *Item Title Template* : `{%1}` ;
- *Item Link Template* : `{%1}` ;
- *Item Content Template* : `{%2}`.

4. Cliquez sur le bouton *Preview* afin d'avoir un aperçu des résultats.

Il ne vous reste plus qu'à récupérer l'adresse du flux RSS qui va être généré et qui sera de ce type : <http://feed43.com/3000482022150474.xml>.

Le seul problème que nous n'avons pas réussi à résoudre (pour le moment) est le fait que certains caractères accentués n'étaient pas correctement restitués. Notez aussi que vous pouvez avoir des balises qui "se baladent" dans le flux RSS mais rien ne vous empêche d'utiliser un Pipe Yahoo! et le module des expressions régulières afin de remplacer une occurrence par "rien" (c'est une manière de la supprimer). Pour de plus amples détails, reportez-vous à notre article du mois dernier. En dépit de ces quelques imperfections, Feed43 reste un très bon service.

Page2RSS

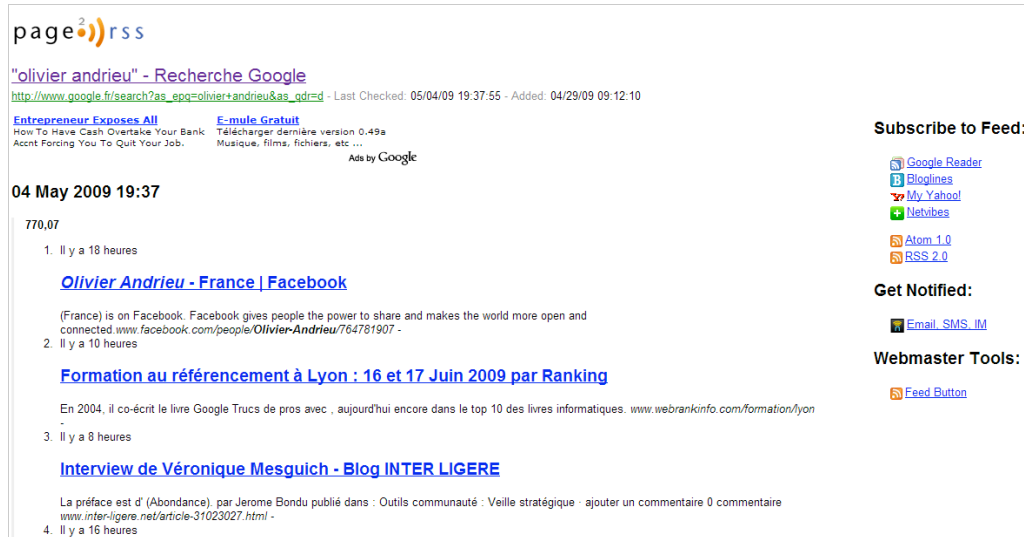
Page2RSS (<http://page2rss.com/>) permet également de générer un flux RSS à partir de n'importe quel type de page web.

1. Saisissez l'adresse URL de la page voulue :

http://www.google.fr/search?as_epq=olivier+andrieu&as_qdr=d.

C'est un autre exemple d'ego-surfing... ;-)

2. Récupérez le lien ou cliquez directement sur un des liens renvoyant vers des agrégateurs en ligne : Google Reader, Bloglines, etc.



page²)rss

"olivier andrieu" - Recherche Google
http://www.google.fr/search?as_epq=olivier+andrieu&as_qdr=d - Last Checked: 05/04/09 19:37:55 - Added: 04/29/09 09:12:10

[Entrepreneur Exposes All](#) [E-mule Gratuit](#)
How To Have Cash Overtake Your Bank Télécharger dernière version 0.49a
Acct Forcing You To Quit Your Job. Musique, films, fichiers, etc ...

Ads by Google

04 May 2009 19:37

770,07

- Il y a 18 heures
[Olivier Andrieu - France | Facebook](#)
(France) is on Facebook. Facebook gives people the power to share and makes the world more open and connected www.facebook.com/people/Olivier-Andrieu/764781907
- Il y a 10 heures
[Formation au référencement à Lyon : 16 et 17 Juin 2009 par Ranking](#)
En 2004, il co-écrit le livre Google Trucs de pros avec . aujourd'hui encore dans le top 10 des livres informatiques. www.webrankinfo.com/formation/lyon
- Il y a 8 heures
[Interview de Véronique Mesguich - Blog INTER LIGERE](#)
La préface est d' (Abondance), par Jerome Bondu publié dans : Outils communauté : Veille stratégique - ajouter un commentaire 0 commentaire www.inter-ligere.net/article-31023027.html
- Il y a 16 heures

Subscribe to Feed:

- Google Reader
- Bloglines
- My Yahoo!
- Netvibes
- Atom 1.0
- RSS 2.0

Get Notified:

- Email
- SMS
- IM

Webmaster Tools:

- Feed Button

Notez qu'il est aussi possible de paramétrer une alerte par e-mail, SMS ou messagerie instantanée. La simplicité de ce service a réellement de quoi surprendre ! De fait, tout va très vite mais vous ne pouvez pas filtrer les modifications qui seront enregistrées et donc pour lesquelles vous allez recevoir une alerte. La page cible est vérifiée une fois par jour.

WatchThatPage

WatchThatPage (<http://www.watchthatpage.com>) est un autre outil de veille qui vous permet de surveiller les changements intervenus sur une page web.

Ce service ne fonctionne pas avec les pages FTP ou celles utilisant les protocoles SSL ou HTTPS. Par ailleurs, il n'est pas possible d'enregistrer les changements apportés à un document PDF, un fichier Word ou Excel. Enfin, vous ne pouvez pas surveiller les pages qui, pour l'instant, n'existent pas (mais qui, à l'avenir, pourront être publiées). Indiquez tout d'abord la fréquence de l'alerte e-mail (tous les jours, certains jours de la semaine).

- Dans la zone de texte *Check for changes at*, indiquez à quelle heure le programme doit se lancer.
- Cliquez sur le bouton *Add update time* si vous souhaitez ajouter d'autres heures auxquelles une vérification sera procédée.
- Dans la liste déroulante *Email content*, précisez si vous voulez que l'e-mail que vous recevrez indique seulement qu'un changement est intervenu ou s'il doit afficher les changements intervenus (*Show me the changes*).
- Cochez la case *Report broken pages* si vous voulez être alerté d'un possible problème survenu lors de la vérification de la page.
- Cliquez sur le bouton *Add keyword matching*.
- Définissez un ou plusieurs mots clés (*Tous les mots, l'expression exacte, un de ces mots, aucun de ces mots*).

Dans la liste déroulante *Match*, deux options sont possibles :

- *The whole page* : ce service tentera de faire coïncider le critère que vous avez défini avec l'ensemble des changements survenus ;

- *Each paragraphe* : dans ce cas, les changements apportés seront indiqués dans des paragraphes distincts et les correspondances possibles seront examinées individuellement.

Account type **Standard**

Advanced Settings...

Email me changes Daily

Weekly on

Halt email updates

Check for changes at

Email content

Report broken pages

Keyword match

Match

With **all** the words

With the exact **phrase**

At least **one** of the words

Without the words

7. Cliquez sur le bouton *Update*.

Notez qu'il est possible de définir une priorité haute en faisant un don de quelques 20 dollars par an. En bref, vous serez servi avant les autres...

8. Dans la fenêtre suivante, cliquez sur le bouton *Add new channel*.

Vous allez définir un canal d'informations.

9. Dans la zone de texte *Add page*, saisissez l'adresse URL de la page à surveiller puis cliquez sur le bouton *Add*.

Notez que vous ne pouvez pas définir des adresses URL générée dynamiquement (mais vous trouverez sans doute des contre-exemples). Du coup, notre test a porté sur cette page qui liste les nouveautés du site Microsoft : <http://www.microsoft.com/fr/fr/>.

L'historique des changements sera visible en cliquant sur le lien *Your changes*.

Une des limitations propres à cet outil est que l'alerte e-mail qui est envoyée ne met pas en évidence les changements intervenus. Vous savez simplement qu'une modification a été apportée sans autre renseignement. De fait, cet outil convient plus à un usage ponctuel et précis. Il existe un Bookmarklet que l'on peut installer à partir de cette page : <http://www.watchthatpage.com/navigationLink.jsp>.

TrackEngine

Vous pouvez vous inscrire au service **TrackEngine** en accédant à cette adresse :

<http://www.trackengine.com/>.

1. Dans la rubrique *Drag-and-drop*, ajoutez le Bookmarklet en le glissant dans la barre des favoris de votre navigateur.

2. Accédez ensuite à la page web que vous souhaitez surveiller.

3. Cliquez sur le bouton *Track me!*

Une fenêtre pop-up va apparaître...

4. Cliquez sur le lien *Expert*.

The screenshot shows the TrackEngine configuration page for tracking a page. The title is "Track this page for jeannoel". The interface is divided into "Basic settings" and "Advanced settings".

Basic settings - / command you, my tireless and faithful servant, to...

- Track this page: (?)
- Give it an easy-to-remember title: (?)
- Please categorize this under folder: (?)
- Or create a new folder: (?)
- How often: (?)
- Please analyze and format: (?)
- And deliver to my mailbox: (?)
- In my favorite color: No highlighting (?)

Advanced settings - / am feeling powerful!

- Include bookmark in AvantGo channel: Yes No (?)
- Send me the report: (?)
- Track changes to: Hyperlinks (?) Images (?) Numbers (?) Dates (?)
- Analyze using: Fuzzy matching (?)
- Check for new content at: (?)
- Disable tracking during: None Weekdays Weekends (?)

Hot stuff - (use commas for more than 1 topic, phrases are accepted)

La liste des options disponibles est la suivante :

- *Track this page* : adresse URL de la page ;
- *Give it an easy-to-remember title* : titre de la page ;
- *Please categorize this under folder* : nom du dossier que vous pouvez créer ;
- *How often* : fréquence d'actualisation (jour, semaine, etc.) ;
- *And deliver to my mailbox* : votre adresse e-mail ;
- *Include bookmark in AvantGo channel* : décidez si oui ou non cette page sera automatiquement ajoutée au canal par défaut (appelé "Avantgo") ;
- *Send me the report* : nombre de changements qui doivent déclencher l'alerte (pas de changement, 1, 2, etc.) ;
- *Track changes to* : types de changements : lien, image, nombre ou date ;
- *Analyze using* : si vous cochez la case *fuzzy matching*, les différences minimales ne seront pas prises en compte ;
- *Check for new content at* : sert à définir une date de départ ;
- *Track my favorite topics* : séparez par des virgules les mots-clés qui doivent déclencher l'alerte (les phrases sont acceptées) ;
- *But don't annoy me with* : saisissez les mots-clés à exclure ;
- *HTTP authentication* : indiquez dans les deux zones de texte les informations d'authentification (nom d'utilisateur et mot de passe) ;
- *Please specify any cookies used* : permet de définir si TrackEngine va utiliser le même cookie que celui qui a servi à télécharger la page ou un autre cookie ;
- *Set as default* : permet de définir les préférences que vous venez de définir comme les options par défaut ;
- *Send confirmation report* : permet d'envoyer un mail de confirmation.

5. Cliquez sur le bouton OK.

6. Cliquez sur le lien *My Bookmarks* afin d'afficher l'ensemble des pages que vous surveillez...

Versionista

Versionista (<http://versionista.com>) est un service d'alerte qui fonctionne à la manière d'un Wiki et qui vous permet de mettre en évidence les changements intervenus sur une page web.

Vous avez droit à 5 pages à surveiller.

1. Ajoutez l'adresse URL à surveiller puis cliquer sur le bouton *Monitor this URL*.

L'adresse URL en vert est celle qui sera surveillée tandis que celles indiquées en violet sont les liens qui ont été détectés comme étant nouveau et qui ne feront pas l'objet d'une surveillance.

2. Ne gardez sélectionnée que celle qui vous intéresse puis cliquez sur le bouton *Monitor checked*.

Le lien *Configuring filtering* permet de définir les éléments qui ne vous intéressent pas :

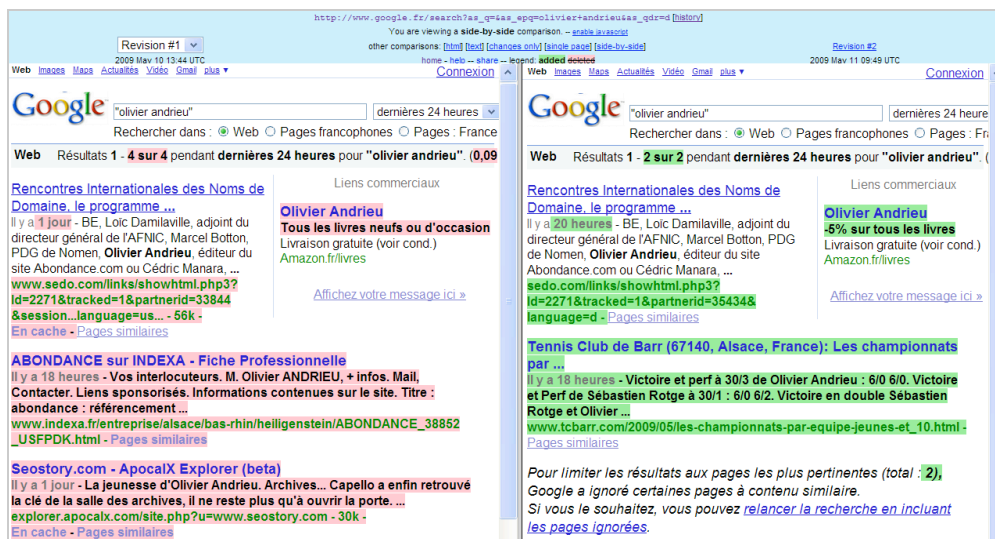
- Les éléments de style CSS, HTML, les codes JavaScripts, etc.
- Les indications chiffrées comme le nombre de visite ;
- La date du jour ;
- L'ensemble des éléments qui font partie du dernier quart de la page Web.

3. Cochez la case *Make this site archive public at* si vous souhaitez rendre public les archives qui ont été générées.

L'adresse URL de la page de surveillance ressemblera à celle-ci :

<http://versionista.com/pub/18290/1/>.

Il vous sera ensuite possible d'effectuer des comparaisons entre les différentes versions qui ont été enregistrées.



Enfin, l'option visible en-dessous vous permet de définir la fréquence de surveillance en fonction de l'offre à laquelle vous avez souscrit.

En haut de la fenêtre, cliquez sur le lien *Your Account* afin de définir les filtres qui seront paramétrés par défaut.

RSSMecano

RSSMecano (<http://rשמecano.com>) permet de créer un flux RSS à partir de n'importe quelle page web...

Il est possible d'essayer gratuitement cette application pour une journée. Sinon, vous pouvez consulter les tarifs à partir de cette adresse : <http://rשמecano.com/produits.php>. Ils sont très raisonnables et, vous l'aurez remarqué, c'est la seule application francophone de notre sélection...

Une fois que vous avez validé votre compte, accédez à votre espace puis cliquez sur le lien *Nouveau*.

Imaginons que nous utilisions cette URL :

http://www.google.fr/search?as_q=jean+noel+anderruthy&hl=fr&as_qdr=d.

Ajoutons maintenant cette adresse URL à RSSMecano. Notez que la seconde méthode fonctionne parfaitement mais il est aussi possible de cocher le bouton radio *Méthode 3*.

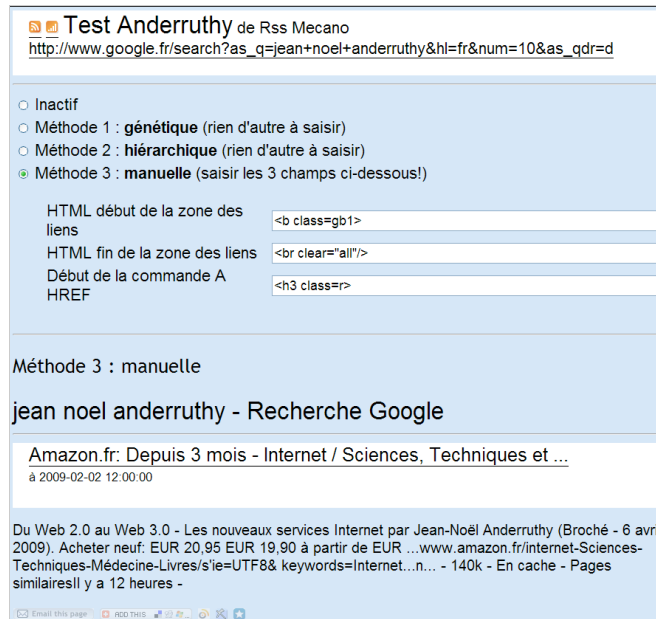
Le code source d'un lien est le suivant :

```
<li class=g><span class="f std">14 avr 2009</span> - <h3 class=r><a href="http://dve4et3.bij.pl/tag,maintenance.html" class=l onmousedown="return
```

clk(this.href,"","'2','")>maintenance</h3><div class="s">Tout savoir sur Maintenance et dépannage d'un PC en réseau, Jean-Noël Anderruthy et les autres nouveautés. Livraison gratuite sur tous les livres en allant. ...
<cite>dve4et3.bij.pl/tag,maintenance.html - </cite>Pages similaires</div>

De fait, le code à utiliser sera celui-ci :

- <b class=gb1> (le début de la page des résultats) ;
- <br clear="all"/> (la fin de la page des résultats) ;
- <h3 class=r> : le début de chaque balise qui annonce un lien.



Il ne vous reste plus qu'à récupérer l'adresse du fil RSS généré par RssMecano, du fil au format mobile HTML pour votre Smartphone, iPhone ou celle du fichier OPML qui contient la liste des fils RSS de votre compte.

Signalons enfin que le créateur de RSS Mecano répond sérieusement et de manière compétente à tous les e-mails.

Dapper

Dapp Factory (<http://www.dapper.net/dapp-factory.jsp>) permet deux choses principalement :

- Générer un flux RSS à partir d'une page web ;
- Personnaliser un flux RSS en fonction de vos besoins.

D'une certaine manière, Dapper est plus simple que Yahoo ! Pipes et offre une interface graphique très efficace. Mais l'utilisation conjointe de ces deux outils vous permet de faire à peu près tout ce qu'il est possible d'imaginer... Nous reprenons ici l'exemple renvoyé par cette adresse URL :

http://www.google.fr/search?as_epq=jean+noel+anderruthy&as_qdr=d&num=100.

1. Dans la liste déroulante *Enter the URL of an RSS Feed or a website that has a feed*, inscrivez l'adresse URL de la page web.
2. Cochez le bouton radio *RSS feed* puis cliquez sur *Next Step*.
Vous allez vous retrouver dans la rubrique *Collect Sample Pages* avec un aperçu de la page web telle qu'elle est visible.
3. Cliquez simplement sur le bouton *Add to basket* afin de l'ajouter à votre panier.
4. Cliquez sur *Next Step*.
Vous allez avoir un message vous signalant que vous pouvez définir d'autres pages web.
5. Validez simplement le fait que vous n'utilisez qu'une seule page source...

De fait, vous pouvez ajouter d'autres pages web afin de créer un flux RSS ou un fichier XML qui regroupe plusieurs requêtes.

6. Là encore, cliquez sur le bouton *Next Step*.

7. Sélectionnez le premier lien qui est visible dans la page des résultats.

L'ensemble de ces mêmes éléments vont apparaître dans la zone de texte *Preview selected content*.

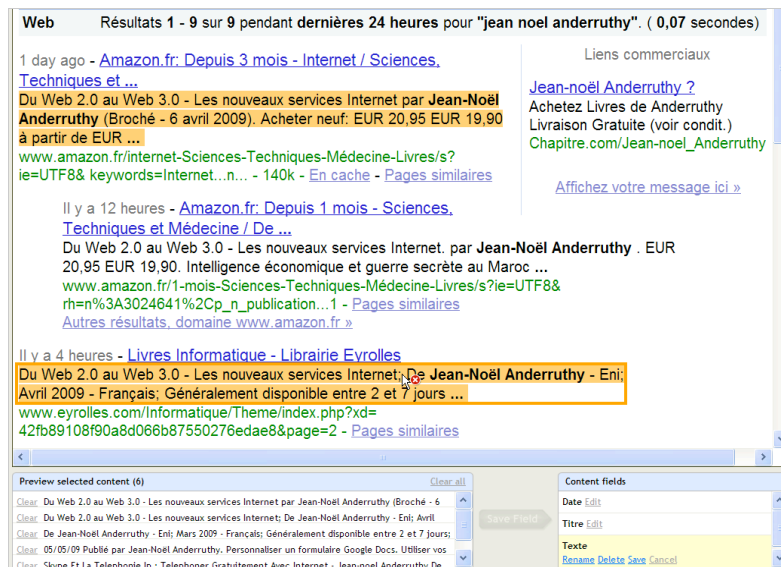
8. Cliquez sur le bouton *Save Field* puis attribuez un nom pour ce groupe ("Titre").

9. Sélectionnez maintenant une indication date ou d'heure puis créer un nouveau champ appelé "Date".

10. Sélectionnez enfin le texte du premier snippet puis créez un nouveau champ appelé "Texte".

11. Cliquez sur le bouton *Next Step*.

12. Dans la zone de texte *Content Fields*, vous allez avoir ces trois champs : Date, Titre et Texte.

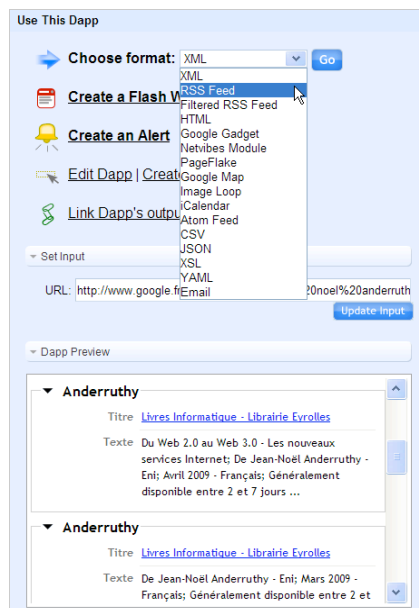


13. Cochez ces trois cases puis cliquez sur le bouton *Save Group*.

14. Attribuez un nom pour votre groupe puis cliquez sur le bouton *Next Step*.

15. Attribuez un nom, une description, un ou plusieurs libellés pour votre *Dapp* puis cliquez sur le bouton *Save*.

16. Dans la liste déroulante *Choose format*, sélectionnez l'option RSS Feed puis cliquez sur le bouton *Go*.



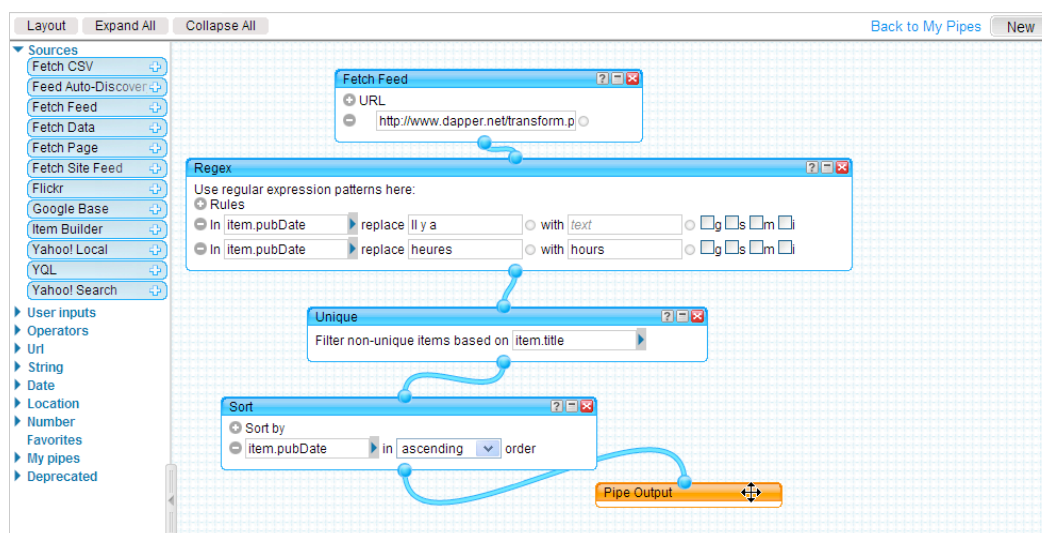
17. Dans la liste déroulante *Item Title*, sélectionnez l'option *Titre*.
 18. Dans la zone de texte *Item Text*, cochez la case *Texte*.
 19. Dans la liste déroulante *Date Published*, sélectionnez l'option *Date*.
 20. Cliquez sur le bouton *Update Input* afin d'avoir un aperçu du résultat final.
- Il ne vous reste plus qu'à récupérer l'adresse URL du flux RSS qui a été généré...
Notez qu'une fois que vous vous êtes inscrit à ce service, vous pouvez vous connecter à votre compte à partir de cette adresse : <http://www.dapper.net/>.

Nous pouvons maintenant parfaire le flux RSS qui est généré en utilisant Yahoo ! Pipes (*voir notre article du mois dernier à ce sujet*).

21. Créer un nouveau Pipe.
22. Ajoutez le module *Fetch Feed*.
23. Dans la zone de texte URL, inscrivez cette récupérée à partir de Dapper ([http://www.dapper.net/transform.php?dapName=EgoAlerte&transformer=RSS&extraArg_title=Titre&extraArg_description\[\]=Texte&extraArg_pubDate=Date&applyToUrl=http%3A%2F%2Fwww.google.fr%2Fsearch%3Fas_epq%3Djean%2520noel%2520anderruthy%26as_qdr%3D%26num%3D100](http://www.dapper.net/transform.php?dapName=EgoAlerte&transformer=RSS&extraArg_title=Titre&extraArg_description[]=Texte&extraArg_pubDate=Date&applyToUrl=http%3A%2F%2Fwww.google.fr%2Fsearch%3Fas_epq%3Djean%2520noel%2520anderruthy%26as_qdr%3D%26num%3D100)).
24. Ajoutez le module *Regex* et ajoutez deux nouvelles règles.
25. Dans la liste déroulante *Rules*, sélectionnez l'option *Item.pubDate*.
26. Remplacez l'occurrence "*Il y a*" par rien (il suffit de laisser la zone de texte suivante vide).
27. Remplacez l'occurrence "*heures*" par ceci : *hours*.

Le problème est que puisque nous avons lancé une recherche à partir de l'interface de Google en français, on obtient ce type d'indication : *Il y a 12 heures*. Après avoir appliqué les expressions régulières, nous aurons ceci : *12 hours*.

28. Ajoutez un module *Unique* afin de supprimer les résultats redondants.
 29. Ajoutez un module *Sort* afin de filtrer les éléments par date.
 30. Dans la liste déroulante *Sort by*, sélectionnez l'option *Item.pubDate* puis, dans la liste déroulante *Order*, sélectionnez l'option *Ascending*.
- Il ne vous reste plus qu'à relier ce dernier module au module *Pipe Output*.



N'oubliez pas de sauvegarder votre Pipe et de récupérer le nouveau flux RSS en cliquant sur le lien *Get as RSS*.

Google XXL: Le plan cadastral avec Google Earth
lundi 4 mai 2009, 14:42:34 →
03/05/09 Publié par Jean-Noël Anderruthy. Le plan cadastral dans Google Earth. Afficher le plan cadastral de votre parcelle dans Google Earth ...

Google XXL
lundi 4 mai 2009, 14:42:34 →
Google SearchWiki en France. Publié par Jean-Noël Anderruthy. Personnaliser les résultats avec Google SearchWiki. Personnaliser les pages de résultats de ...

Google XXL: Une carte Google avec Scribble Maps
dimanche 3 mai 2009, 12:20:28 →
02/05/09 Publié par Jean-Noël Anderruthy. Créer une carte Google Maps avec Scribble Maps. Tracer des cercles Google Maps, ajouter du texte, ...

Livres Informatique personnelle - Librairie Eyrolles
dimanche 3 mai 2009, 12:20:28 →
Maintenance et dépannage d'un PC en réseau; De Jean-Noël Anderruthy - Eni; Janvier 2008 - Français; Généralement disponible entre 2 et 7 jours ...

Affichage 24 / 24

Tout 24

Trier par :
Date
Titre

[Afficher les propriétés du flux...](#)

La boucle est bouclée !

De réels outils de veille ?

Par rapport à une plateforme professionnelle, il est sûr que les limitations des outils présentés dans cet article sont très claires :

- Les délais de surveillance ne vous permettront pas de réagir en temps (presque) réel ;
- Vous n'avez pas de données statistiques des résultats obtenus ;
- Vous ne pouvez pas traiter des sources qui seraient en nombre très important ;
- On a du mal à discerner un quelconque modèle économique pour certains de ces outils ;
- Chaque outil présente des faiblesses et des qualités qui leur sont propres.

Néanmoins et même en version gratuite, ces applications, chacune à leur manière, peuvent constituer un appoint précieux à un processus de veille plus général qui reposera principalement sur les flux RSS "officiels". Il vous suffit de faire votre choix en fonction des besoins de votre société ou de votre activité. Au final, ils vous serviront pour toutes sortes d'usages qui peuvent aller de l'ego-alerte à la veille concurrentielle et ce pour une large gamme d'activités : documentaliste, référenceur professionnel, chef d'entreprise, etc.

Jean-Noël Anderruthy, *webmaster spécialisé dans les technologies Google.*

Réagissez à cet article sur le blog des abonnés d'Abondance :

<http://abonnes.abondance.com/blogpro/2009/05/10-outils-de-surveillance-de-sites-web.html>