

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

L'omniprésence des réseaux sociaux dans notre quotidien fait que l'indexation en temps réel, des informations publiées en ligne, devient un enjeu majeur en termes de stratégie "SEO". Aussi, il est important de "ping" les moteurs de recherche pour leur annoncer la publication d'un nouveau contenu et l'utilisation de services comme PubSubhubbub devient le levier indispensable d'une prochaine maxime du référencement : "aussitôt publié, aussitôt indexé" !!!...

Le temps réel permet de rapprocher les univers virtuels de façon à, justement, élargir notre champ de connaissances et d'informations. Le temps devient alors un facteur de pertinence et d'optimisation de votre site web : plus vous ferez court et plus vous aurez raison. Google, qui est champion de la vitesse dans tous les secteurs qui touchent au Web, vient justement de mettre au point un protocole qui rend l'indexation en temps réel à la portée de chacun. Voyons comment vous pouvez vous y prendre pour mettre un pied dans l'instantanéité...

L'indexation sur les chapeaux de roue

Il n'y a pas si longtemps que cela, les propriétaires de sites web devaient, par exemple, soumettre un plan Sitemap à Google et attendre patiemment que le "GoogleBot" crawle les pages indiquées pour, enfin, voir le contenu de leur site web indexé.

Peu après, les flux RSS ont fait leur apparition et il était alors possible de s'en servir pour générer automatiquement un plan Sitemap "dynamique" des derniers articles parus sur un site.

Oui ! Mais le souci restait lié au fait que ce processus supposait un nombre conséquent d'allers-retours et des infrastructures serveur extrêmement importantes... Que seuls des mastodontes comme Google pouvaient s'offrir.

Par ailleurs, il existait un temps d'attente entre la publication d'un article et son indexation dans les moteurs de recherche.

Le développement des sites de réseautage social et l'essor des services de LiveStreaming et de PlaceStreaming ont rendu nécessaire l'élaboration de plateformes rendant possibles un Web "en temps réel".

Trois ingénieurs de Google (Mihai Parparita, Brett Slatkin et Brad Fitzpatrick) se sont ainsi attaqués à ce problème et ont proposé à la communauté, un nouveau protocole appelé PubSubhubbub. Son objectif est simple à comprendre (pas comme son nom à retenir...) : raccourcir le temps virtuel qui sépare les sites web des internautes jusqu'à le rendre proche de zéro. En bref : aussitôt publié, aussitôt indexé, aussitôt lu...

Les avantages de PubSubhubbub

PubSubhubbub signifie "server-to-server web-hook-based pubsub" (soit publish/subscribe nom du protocole). L'abréviation couramment utilisée est celle-ci : PuSH.

Le "Working Draft" de la "bête" est visible à cette adresse :

<http://pubsubhubbub.googlecode.com/svn/trunk/pubsubhubbub-core-0.3.html> et le fichier d'aide à partir de celle-ci : <http://code.google.com/p/pubsubhubbub/>. Il existe de nombreux exemples qui vous permettent d'implémenter rapidement ce même protocole que vous soyez "souscripteur" ou "éditeur".

Les principales caractéristiques sont :

- Une mise en place d'un protocole de publication ou de souscription ;
- Les fils d'actualités deviennent des flux en lecture continue (Streams) ;
- Une réduction des temps de latence.

Le protocole PubSubhubbub suit ce cheminement :

- Un éditeur de contenu propose un flux RSS ;
- Ce flux RSS fait mention du serveur hub (concentrateur) dans les lignes de déclaration du fichier XML (*Extensible Markup Language*) ou ATOM (format de syndication basé sur XML) ;
- Un serveur (l'abonné) parcourt normalement le fichier concerné ;
- Si le fichier de flux déclare l'adresse du hub qu'il utilise, le serveur abonné peut alors souscrire au flux diffusé par le hub ;
- Dès que l'éditeur a informé le hub qu'une mise à jour est disponible, ce dernier récupère l'élément nouveau et diffuse les changements aux différents serveurs concernés : *"J'ai fini le boulot ; tu n'as plus qu'à venir te servir !"*

Le hub fonctionne donc comme un intermédiaire entre les éditeurs de contenu et les abonnés. Et c'est seulement lui qui gère la publication, le processus d'abonnement (ou de désabonnement) ainsi que l'acheminement des messages. L'ensemble de ce processus (en cascade et qui supprime les commandes doublons) se déroule en quelques secondes...

Autre particularité importante : dans un processus normal, c'est l'abonné qui interroge le serveur (en mode Pull) afin de savoir si une mise à jour est disponible alors que là, c'est le contraire : le hub lance une alerte en mode Push. Par exemple, **Ping-O-Matic** (<http://pingomatic.com>) est un service qui alerte des agrégateurs, comme FeedBurner ou NewsGator, que du nouveau contenu a été publié et qu'ils peuvent interroger votre flux RSS.

The screenshot shows the Ping-O-Matic website interface. At the top, there is a navigation bar with links for Home, Stats, and Blog. Below this, a section titled "PINGING SEARCH ENGINES" explains the service: "What is this? Ping-O-Matic is a service to update different search engines that your blog has updated." A "Learn more..." link is provided.

The main content area is titled "Blog Details" and contains three input fields:

- Blog Name:
- Blog Home Page:
- RSS URL (optional):

Below the form is a section titled "Services to Ping" with two columns of checkboxes:

Common Services (Uncheck all)	Specialized Services
<input checked="" type="checkbox"/> Weblogs.com	<input type="checkbox"/> Audio.Weblogs
<input checked="" type="checkbox"/> Syndic8	<input type="checkbox"/> RubHub
<input checked="" type="checkbox"/> PubSub.com	<input type="checkbox"/> GeoURL
<input checked="" type="checkbox"/> Moreover	<input type="checkbox"/> A2B GeoLocation
<input checked="" type="checkbox"/> News Is Free	<input type="checkbox"/> BlogShares
<input checked="" type="checkbox"/> Spinn3r	
<input checked="" type="checkbox"/> SkyGrid	
<input checked="" type="checkbox"/> Blo.gs	
<input checked="" type="checkbox"/> NewsGator	
<input checked="" type="checkbox"/> Blogdigger	
<input checked="" type="checkbox"/> Weblogalot	
<input checked="" type="checkbox"/> Topic Exchange	
<input checked="" type="checkbox"/> Bloglines	
<input checked="" type="checkbox"/> Collecta	
<input checked="" type="checkbox"/> Feed Burner	
<input checked="" type="checkbox"/> My Yahoo!	
<input checked="" type="checkbox"/> BlogStreet	
<input checked="" type="checkbox"/> Icerocket	
<input checked="" type="checkbox"/> Google Blog Search	
<input checked="" type="checkbox"/> PostRank	
<input checked="" type="checkbox"/> Superfeedr	

At the bottom of the form is a large button labeled "Send Pings »" with a mouse cursor icon pointing to it.

La difficulté est que les services concernés doivent alors visiter l'adresse URL du flux afin d'en récupérer le contenu frais. En bref, vous pingez ces services et ils viennent "butiner" votre contenu.

Avec PubSubhubbub, le hub choisi agrège l'élément mis à jour et le diffuse en mode multidiffusion. La différence est de taille en termes de rapidité et d'économie de bande passante pour votre site.

Une vidéo montre la rapidité d'indexation entre une mise à jour faite par un utilisateur et le flux d'actualités qui sera visible : <http://www.youtube.com/watch?v=P-9KoVwQFVQ&feature=related>.

On peut citer ces quelques services Google (et autres) utilisant PuSH en tant que souscripteurs : Google Reader, Blogger, Google Buzz, FeedBurner, les alertes Google, Friendfeed (<http://friendfeed.com>), Ping.fm (<http://ping.fm>), Netvibes, Status.net (<http://status.net>), etc.

Ces services l'utilisent déjà en tant qu'éditeurs : Typepad, Wordpress, Posterous (<http://www.tumblr.com>), Tumblr (<http://www.tumblr.com>), etc.

Un type d'utilisation en mode "Application", l'**API Feed Ajax de Google** (<http://code.google.com/apis/ajaxfeeds/>), interroge PubSubhubbub pour mettre à jour les Widgets concernés, et ce en temps réel (ou presque). C'est ce qui explique qu'entre le moment où vous publiez un article dans Blogger ou FriendFeed et l'instant où il est visible dans Google Reader, il peut se passer moins d'une minute montre en main.

Notez aussi qu'il existe d'autres avantages et, notamment, comparés à un service très similaire appelé RSSCloud. Google a eu la bonne idée de publier un tableau visible à cette adresse : <http://code.google.com/p/pubsubhubbub/wiki/ComparingProtocols>. Les différents protocoles sont examinés selon leurs performances en termes de sécurité, temps de latence, consommation de bande passante, difficulté d'implémentation... (*light pings vs fat pings !*).

Mise en place de PubSubhubbub

Comment déclarer un hub dans un flux RSS ou ATOM ? Il vous suffit d'ajouter cette ligne dans le fil d'actualité :

```
<?xml version="1.0"?>
<rss xmlns:atom="http://www.w3.org/2005/Atom">
  <channel>
    <atom:link rel="hub" href="L'adresse du hub voulu"/>
    ...
  </channel>
</rss>
```

La liste des clients qu'il est possible d'utiliser (dont un Bookmarklet pour les CMS qui ne possèdent pas de fonction de Ping automatique) est visible sur cette page :

<http://code.google.com/p/pubsubhubbub/wiki/PublisherClients>.

Les adresses des hubs sont disponibles sur cette page :

<http://code.google.com/p/pubsubhubbub/wiki/hubs>.

C'est aussi un des points qui font de PuSH un service intéressant : il n'est pas contrôlé par Google et n'importe quelle entité a la possibilité de créer son propre hub... Le processus de développement de ce protocole est complètement transparent...

Par exemple, **Superfeedr** (<http://superfeedr.com>) fonctionnera comme "hub parent" sur lequel vous pouvez créer votre propre service. **BackType** (<http://www.backtype.com>) utilise ce site pour créer son propre "hub enfant". Les détails de l'implémentation sont expliqués sur cette page : <http://tech.backtype.com/pubsubhubbub-publishing>.

Notez que Backtype sert à mesurer l'impact social de votre entreprise et que les outils accessibles via les "Backtype Labs" proposent des fonctionnalités permettant d'afficher le nombre de tweets qui font un lien vers une page de votre site.

The screenshot shows the 'backtweets' search interface. At the top, there are tabs for 'URL Search' and 'Advanced Search'. The search bar contains the URL 'http://actu.abondance.com/2010/07/twitter-2eme-moteur-de-recherche.html'. Below the search bar, the results are displayed as a list of tweets. Each tweet includes a profile picture, the user's name, the tweet text, and the time it was posted. The tweets are all retweets of a post from @martinlessard, which mentioned 'Twitter 2ème moteur de recherche mondial' and provided a link to the article on Abondance.com.

À partir de là, il existe plusieurs cas de figure :

- Soit vous êtes déjà sur une plate-forme qui intègre cette technologie et vous n'avez rien à faire (ou presque)...
- Soit, en fonction du CMS utilisé, vous pouvez installer un module complémentaire et le tour est joué ;
- Soit vous devez mettre la main dans le cambouis et faire un peu de programmation afin de pouvoir *ping* le hub de votre choix.

Rappelez-vous que vous pouvez utiliser ce protocole en tant qu'abonné (et donc en vue d'accélérer vos systèmes d'information) ou en tant qu'éditeur.

Vous trouverez un outil de publication et de diagnostic disponible à cette adresse : <https://pubsubhubbub.appspot.com/publish> mais celui qui est proposé sur cette page est beaucoup plus parlant : <http://www.grailsworks.com/grailsbox/pubSubhubBub> (avec une excellente explication sur le sujet : <http://www.odelia-technologies.com/PubSubhubbub>).

Pubsubhubbub et Wordpress

Un plugin Wordpress est disponible à partir de cette page : <http://wordpress.org/extend/plugins/pushpress/>. Il vous suffit simplement d'installer et activer ce plugin. La différence avec les autres plugins Wordpress est qu'il intègre déjà un hub PuSH.

PubSubhubbub (<http://wordpress.org/extend/plugins/pubsubhubbub/>) est un autre plugin Wordpress qui nécessite que vous indiquiez quels sont les hubs que vous souhaitez pinger. Par défaut, l'adresse du hub de démonstration de Google App Engine (<http://code.google.com/appengine/>) et celui de Superfeedr (<http://superfeedr.com>) sont déclarés.

WP Pubsubhubbub (<http://wordpress.org/extend/plugins/wp-pubsubhubbub/>) remplit exactement les mêmes fonctions...

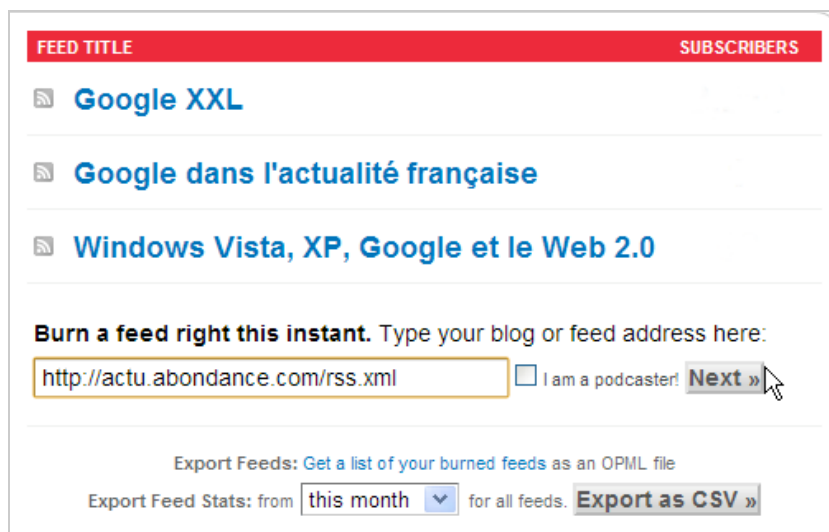
Une fois le plugin installé, accédez à l'interface d'administration de Wordpress (*Réglages – Écriture*) puis rajoutez, dans la rubrique *Services de mise à jour*, le ou les hubs voulus. Il est conseillé d'en spécifier au moins deux en cas de défaillance de l'un ou de l'autre.

Pour information, une liste des services de Ping implémentés dans Wordpress est accessible à partir de cette page : <http://www.instant-info-online.com/wordpress-compressed-all-inclusive-ping-list.html>.

L'indexation en temps réel avec FeedBurner

Ce service permet de profiter, pour ainsi dire, du beurre et de l'argent du beurre...

1. Muni de vos identifiants Google, accédez à FeedBurner : <http://feedburner.google.com/fb/a/myfeeds>.
2. Dans la zone de texte *Burn a feed right this instant*, indiquez l'adresse de votre flux puis cliquez sur le bouton *Next*.



The screenshot shows the FeedBurner interface. At the top, there are two tabs: "FEED TITLE" (active) and "SUBSCRIBERS". Below the tabs, there is a list of feeds:

- Google XXL
- Google dans l'actualité française
- Windows Vista, XP, Google et le Web 2.0

Below the list, there is a section titled "Burn a feed right this instant. Type your blog or feed address here:". It contains a text input field with the URL "http://actu.abondance.com/rss.xml", a checkbox labeled "I am a podcaster", and a "Next »" button.

At the bottom, there are two options for exporting feeds:

- Export Feeds: Get a list of your burned feeds as an OPML file
- Export Feed Stats: from a dropdown menu set to "this month" for all feeds. An "Export as CSV »" button is next to it.

3. Indiquez un titre puis modifiez éventuellement l'adresse qui est indiquée. Par défaut, deux services sont automatiquement activés : *BrowserFriendly* et *FeedBurner Stats*.
4. Cliquez sur le bouton *Next*.

Si vous souhaitez obtenir des statistiques précises de votre fil d'actualités, cochez les trois cases visibles puis cliquez sur le bouton *Next*.

La page qui est visible permet de faciliter la transition entre les flux publiés par défaut par votre plate-forme de publication de blogs et celle de FeedBurner.

Dans le cas contraire, vous devez modifier l'adresse du flux RSS telle qu'elle est visible sur votre site pour indiquer l'adresse du flux "redirigé".

1. Une fois que vous avez "recuit" le flux RSS de départ dans FeedBurner, accédez à ses paramètres puis cliquez sur l'onglet *Publicize*.
2. Cliquez ensuite sur le lien *PingShot* puis sur le bouton *Activate*.

Abondance Google feedburner

Edit Feed Details... | Delete Feed... | Transfer Feed...

Analyze Optimize **Publicize** Monetize Troubleshootize My Feeds

↓ SERVICES

Headline Animator
Display rotating headlines

BuzzBoost
Republish your feed as HTML

Email Subscriptions
Offer feed updates via email

PingShot
Notify services when you post

FeedCount
Show off your feed circulation

PingShot
Most web-based feed reading services will check for updates on their own time. Give 'em a push with PingShot.

Notify interested services when my feed changes

Activate This service is inactive

PingShot est un répertoire ouvert auquel tout le monde peut participer en utilisant le protocole PubSubhubbub.

Nous n'avons pas réussi à trouver la liste des services fédérés mais, de mémoire, voici les plus connus : Technorati, My Yahoo!, Bloglines, Pingomatic, Newsgator, Google Blog Search Pinging Service, Icerocket, Weblogs.com, Blogdigger.

En accédant au code source du flux RSS (*Optimize - XML Source*), vous pourrez découvrir ces quelques lignes de code :

```
<atom10:link xmlns:atom10="http://www.w3.org/2005/Atom" rel="self"
type="application/atom+xml" href="http://feeds.feedburner.com/abondance" />
<feedburner:info uri="abondance" />
<atom10:link xmlns:atom10="http://www.w3.org/2005/Atom" rel="hub"
href="http://pubsubhubbub.appspot.com/" />
```

Bien entendu, vous pouvez aussi activer les autres briques qui sont proposés par FeedBurner :

1. Cliquez sur l'onglet *Publicize* puis le lien *Socialize*.
2. Ajoutez votre compte Twitter puis définissez les options de formatage voulues :
 - *Post content* : titre ou/et description ;
 - *Leave room for retweets* : laissez de la place de telle façon que votre Tweet peut être retweeté en ajoutant un commentaire ;
 - Ajout des Hashtags générés automatiquement ou non ;
 - Ajout d'un texte additionnel (comme un Hashtag mentionnant votre marque) ;
 - *Item limit* : nombre d'éléments visibles ;
 - *Keyword filter* : mots-clés de filtrage.

La fenêtre *Sample Item Preview* s'actualise au fur et à mesure des choix que vous avez fixés.

Add a Twitter account

Formatting Options

Post content:

Include link

Leave room for retweets

Hash tags:

Use inline hash tags

Additional text: Add at the

Item Selection

Item limit: Post up to new items per feed update

Item order: Order items by

Keyword filter:

e.g. toast, purple cow

Only post items containing a keyword in

Sample Item Preview

#abondance Twitter 2ème moteur de recherche mondial avec 800 millions de recherches par jour ! <http://goo.gl/fb/abcd>

Les éléments qui composent votre fil d'actualités seront automatiquement repris sur votre profil Twitter.

Notez qu'un service comme TwitterFeed nous paraît aussi intéressant puisqu'il vous permet de spécifier un service de "raccourcisseur d'URL" comme Bit.ly (<http://bit.ly>) et, en conséquence, le nombre de clics générés sur les adresses URLs "compressées" par vos soins.

Step 1: Create Feed Step 2: Configure Publishing Services Step 3: Done

Step 1: Name Feed & Add Source URL

Feed Name

RSS Feed URL

Active

Advanced Settings

Update Frequency
Check for new posts And post up to new update(s) at a time.

Post Content
Include Post Link
Shorten link through

Post Sorting
Post new items based on

Post Prefix
Prefix each tweet with: (max. 20 characters)

Post Suffix
Suffix each tweet with: (max. 20 characters)

Keyword Filter
Separate multiple words with a space (for example, "apple orange" will match any posts that contain the term "apple" OR "orange")
 Filter your posts by using keywords to auto-approve new posts.

Réseaux sociaux et recherche en temps réel

Nous avons beaucoup parlé dans cet article de Google et, d'une certaine manière, d'un service qui appartient aussi à cette société mais, peut-être, que les réels enjeux sont ailleurs... Et ce même si ce qui prévaut aux États-Unis n'est pas comparable avec la situation française. Le blog d'Abondance vient de reprendre un chiffre intéressant (<http://actu.abondance.com/2010/07/twitter-2eme-moteur-de-recherche.html>) : Twitter se positionnerait actuellement comme le deuxième moteur de recherche mondial avec 800 millions de requêtes par jour, soit 24 milliards par mois, derrière Google mais devant Yahoo! et Bing...

Ces statistiques sont à relativiser : un grand nombre de recherches sont effectuées depuis des clients externes, et ce en utilisant l'API de Twitter : Tweetdeck, Seismic, etc. Il faut aussi noter que ce chiffre englobe des requêtes automatisées puisqu'un grand nombre d'internautes utilisent Twitter comme un service de veille.

Par ailleurs, selon l'institut Hitwise, Facebook a représenté, en mars 2010, 7,07% de l'ensemble du trafic internet américain, contre 7,03% pour Google...

L'utilisation de ces plates-formes de conversation sociale tendent aussi vers le "Search" et donc à remplacer les moteurs "traditionnels". Nous ne sommes pas loin de penser également que le développement du Web mobile ne fera qu'accentuer cette tendance : l'information se sociabilise...

La recherche en temps réel est donc un objectif en termes de SEO puisque plus vous serez rapidement indexés, plus votre contenu aura des chances d'être "viralisé" et, en conséquence, popularisé. Et c'est un des paradoxes de la recherche en temps réel : l'indexation de votre contenu n'est plus le fait des seuls robots mais aussi celui des internautes.

Expliquons-nous : avant seuls existaient des systèmes de publication de contenu. Puis, sont arrivés des systèmes performants d'agrégation de contenu qui fonctionnent aussi comme des services de génération de contenu assisté : les Widgets en sont un bon exemple. L'indexation en temps réel permet de synchroniser ces brides de site web qui fera que, par exemple, le contenu d'un blog s'affichera aussi sur la page personnalisée de tel ou tel internaute.

Si le Ranking de votre page dépend de sa popularité, non plus seulement au travers du Web (le PageRank) mais aussi auprès des internautes, l'indexation en temps réel devient un facteur critique sur lequel vous devez porter vos efforts.

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

Réagissez à cet article sur le blog des abonnés d'Abondance :
<http://blog-abonnes.abondance.com/2010/07/pubsubhubbub-ping-et-indexation-en.html>