

La recherche temps réel avec Twitter : outils et perspectives (2ème partie)

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Le temps réel est à la mode. On parle de ce mode de recherche, et de son site emblématique, Twitter, comme d'un potentiel concurrent de Google. Qu'en est-il exactement ? Twitter peut-il réellement nous aider à mieux chercher qu'avec les moteurs "classiques" ? Quels sont les outils rattachés à Twitter qui permettent d'effectuer ce nouveau type de recherche ? Voici quelques pistes de réponse et de réflexion, ainsi que de nombreux outils à tester pour vous faire votre propre opinion. Dans cette deuxième partie, nous allons plus particulièrement nous attacher à décrire les services que proposent Google et ses concurrents pour effectuer des recherches sur des documents récents, ainsi que de nombreux outils en ligne qui ont été spécifiquement créés pour cette tâche...

Nous continuons ce mois-ci notre exploration des outils permettant de rechercher de l'information en temps réel avec Twitter. Après les outils proposés par Twitter lui-même ou les services spécialisés créés pour l'occasion, nous décrivons ce mois-ci d'autres possibilités autour de Google, Yahoo!, Bing, et bien d'autres moteurs de recherche qui démontrent l'incroyable richesse et le potentiel de cette nouvelle façon de rechercher l'information sur le Web !

Google

Vous savez déjà que Google est le seul moteur de recherche à proposer des résultats au jour le jour : cliquez sur le lien *Recherche avancée* puis servez-vous de la liste déroulante *Date* et sélectionnez cette option : *Dernières 24 heures*. Vous pouvez ensuite manipuler l'adresse URL pour n'afficher que les résultats des 5 derniers jours et ce, en modifiant le commutateur [&as_qdr=d](#) par celui-ci : [&as_qdr=d5](#). Notez qu'à partir de l'interface classique, vous pouvez aussi utiliser ce commutateur : [&as_qdr=h](#). Mais le nombre de résultats renvoyés est beaucoup plus restreint et nous ne sommes pas sûrs que ce soit finalement la bonne méthode...

La version anglaise vous permet d'accéder à une interface plus poussée quand, après une recherche, vous cliquez sur le lien *Show options*.

Nous avons eu l'impression que le lien *Recent results* permettait d'afficher les liens les plus récents mais classés en fonction de leur pertinence. Quoiqu'il en soit vous pouvez les trier par date en cliquant sur le lien *Sort by date* et, finalement, ne retenir que les résultats des dernières 24 heures en cliquant sur le lien correspondant (*Past 24 hours*).

Enfin, et si vous souhaitez filtrer les résultats à une langue en particulier, cliquez sur le lien *Advanced Search* puis, dans la liste déroulante *Language*, sélectionnez l'option *French*. Cliquez sur le bouton *Advanced Search* puis, de nouveau, sur le lien *Sort by date*.

On obtient alors une adresse URL qui ressemble à celle-ci :

[http://www.google.com/search?q="google%20wave"&hl=en&lr=lang_fr&sa=N&tbo=1&tbs=qdr:d,sbd:1](http://www.google.com/search?q=). Nous avons lancé une recherche sur l'expression "Google Wave" mais vous pouvez aussi bien filtrer plus finement les résultats en utilisant cette requête : [intitle:"google wave"](#) et en forçant l'affichage de 100 liens par page

([http://www.google.com/search?hl=en&tbo=1&tbs=qdr:d,sbd:1&q=intitle:"google+wave"&lr=lang_fr&num=100](http://www.google.com/search?hl=en&tbo=1&tbs=qdr:d,sbd:1&q=intitle:)).

The screenshot shows a Google search interface. The search bar contains the text "intitle:google wave". To the right of the search bar are links for "Advanced Search" and "Preferences". Below the search bar, there are radio buttons for "Search the Web" (selected) and "Search French pages". The search results are displayed in a table-like format. The left column contains filters and sorting options: "All results", "Videos", "Forums", "Reviews", "Any time", "Recent results", "Past 24 hours", "Past week", "Past year", "Sort by relevance", "Sort by date", "Standard results", "Images from the page", and "More text". The right column displays the search results. The first result is titled "google wave : Inovia Blog" and includes a link to "Translate this page". The second result is titled "Google Wave : la révolution n'est peut être pas celle que l'on ..." and also includes a link to "Translate this page". The third result is titled "Premier avis sur Google Wave / Entreprise Marketing" and includes a link to "Translate this page".

Filters	Results
› All results	google wave : Inovia Blog - [Translate this page]
Videos	2 hours ago - La récente conférence Google I/O a été marquée par la disponibilité prochaine de Google Wave, une étonnante plateforme de communication et de collaboration ...
Forums	blog.inovia-conseil.fr/?cat=22 - Similar pages
Reviews	Google Wave : la révolution n'est peut être pas celle que l'on ... - [Translate this page]
Any time	3 hours ago - Par Kadda SAHNINE. Publié le juin 5, 2009 . Logo Wave La récente conférence Google I/O a été marquée par la disponibilité prochaine de Google Wave, ...
Recent results	blog.inovia-conseil.fr/?p=38 - Similar pages
› Past 24 hours	Premier avis sur Google Wave / Entreprise Marketing - [Translate this page]
Past week	5 hours ago - Les Meilleurs Liens Marketing ! (BETA). Le Top A venir Ajouter un lien .
Past year	Accueil · Login · S'enregistrer · Le Blog · Déconnexion ...
Sort by relevance	www.conseilsmarketing.fr/entreprise-marketing/story.php?title=premier-avis-sur-google-wave - Similar pages
› Sort by date	
Standard results	
Images from the page	
More text	


Il peut être intéressant d'afficher les *Backlinks* d'un site en utilisant ce type de requête : [link:abondance.com](#). Cette possibilité fonctionne maintenant en cliquant, par exemple, sur les liens *Recent results* et *Sort by date*. L'adresse URL sera alors de ce type : <http://www.google.com/search?hl=en&tbo=1&tbs=qdr:w,sbd:1&q=link:abondance.com&aq=f&oq=&aqi=>. Le seul souci est que les résultats ne sont pas limités aux pages les plus récentes mais englobent l'ensemble des liens détectés (qui, rappelons-le, n'est pas exhaustif avec la requête [link:](#) sur Google). Par ailleurs, cette astuce fonctionne avec certains noms de domaine et pas avec d'autres. Cela dit, rien ne vous empêche alors d'utiliser un programme de *Web scrapping* pour générer un flux RSS.

Le problème ne se pose pas si vous lancez ce type de requête : [site:abondance.com](#). C'est une bonne manière de voir quelles sont les pages qui ont été publiées récemment, même si le site cible ne propose pas de flux RSS ou que certains de ces éléments restent "muets".

Bing contre Google


On s'est vite posé la question du moteur qui intégrera en premier les données de Twitter dans ses résultats classiques. Pour une fois, c'est Bing qui a gagné puisque vous pouvez, à partir de la version américaine du moteur, explorer certaines données issues de Twitter.

1. Accédez à Bing (<http://www.bing.com/>) puis cliquez sur les liens *France* et *United States - English*.
2. Testez les résultats renvoyés par ces requêtes : [mashable Twitter](#) ou [matt cutts tweets](#) ([#mattcutts tweet](#) fonctionne aussi) ou encore [@techcrunch](#).

mashable Twitter 

ALL RESULTS 1-10 of 137,000 results · [Advanced](#)

[Pete Cashmore](#) on Twitter [@mashable](#)

 Facebook's '09 Revenue to Top \$500 Million - <http://bit.ly/2JZslg> · 9 minutes ago

Twitter keeps giving me the Fail Whale. Off to play the Die Fail Whale Game :) <http://bit.ly/UQrT0> · 20 hours ago

[See more tweets](#) · [Source: twitter.com](#)

[Twitter Lists](#)

Twitter Resource Lists ... **Twitter** Brand Sponsors. These brands would like to engage with our social media community on **Twitter**. mashable.com/category/twitter-lists · [Cached page](#)

[Twitter](#)


Twitter Brand Sponsors. These brands would like to engage with our social media community on **Twitter**. If you like them, Tweet them! mashable.com/tag/twitter · [Cached page](#)

[Pete Cashmore \(mashable\) on Twitter](#)

The hottest **Twitter** news, **Twitter** tips and **Twitter** help. Plus, the best social media links around! <https://twitter.com/mashable> · [Cached page](#)

Ce ne sont pas, à proprement parler, des résultats "en temps réel" mais plutôt une série de lien de type "réseau sociaux" comme les profils Facebook que l'on peut voir dans les résultats des moteurs. Du coup, il faut un peu aider le moteur de recherche en lui indiquant la route à suivre comme pour cette requête : [tour de france twitter](#). Mais, dans ce cas, tous les moteurs répondent présents. Ici Yahoo! :

[Tour De France Twitter - News Results](#)

 [Twitter De France](#) Deseret News - Jul 04 09:21pm
A sample of tweets posted Saturday by **Tour de France** riders:

[Tour de France \(tourfrance\) on Twitter](#)

Tour de France news and info. ... Valverde out of **Tour de France** (AP) <http://tinyurl.com/nllxv6> 2:56 PM Jun 23rd from twitterfeed ... twitter.com/tourfrance - [Cached](#)

[Frank Steele \(TdFblog\) on Twitter](#)

Tour de France ... **Twitter** is a free service that lets you keep in touch with ... Bio **Tour de France** blog updates. 16 Following. 490 Followers. 2, ... twitter.com/TdFblog - [Cached](#)

Ou Google :

 [Recherche avancée](#)
[Préférences](#)

Rechercher dans : ☒ Web ☐ Pages francophones ☐ Pages : France

Web Résultats 1 - 10 sur un total d'environ 22 900 000 pour **tour**

[Tour de France \(tourfrance\) on Twitter](#) - [[Traduire cette page](#)]
BRIEF-Cycling-Astana name Armstrong in **Tour de France** team (Reuters) http://tinyurl.com/ml259f1:48 AM Jun 19th from twitterfeed ...
[twitter.com/tourfrance](#) - [En cache](#) - [Pages similaires](#)

[Frank Steele \(TdFblog\) on Twitter](#) - [[Traduire cette page](#)]
Tour de France race tweets and links. ... **Twitter** is a free service that lets you keep in touch with people through the exchange of quick, frequent answers ...
[twitter.com/TdFblog](#) - [En cache](#) - [Pages similaires](#)
[Autres résultats. domaine twitter.com »](#)

De nombreux commentateurs ont signalé que les messages Twitter seront accessibles via Google News, dans sa nouvelle version (nom de code Flipper d'après cet article : <http://blogs.lexpress.fr/virtuel/2009/07/exclu-bientot-les-resultats-tw.php>). Le porte-parole de Google s'en explique de cette façon : *si le président des Etats-Unis décide de faire une annonce importante en utilisant Twitter, on peut légitimement espérer que ce même message soit visible dans Google Actualités et, en conséquence, dans les Serps (Search Engine Result Pages) du moteur classique*. C'est un premier pas intéressant dans l'indexation du contenu des réseaux sociaux dans les moteurs de recherche "non spécialisés".

L'actualité sportive en 4 résultats dédiés

On peut penser que c'est réellement l'exemple parfait d'une recherche qui nécessite que les résultats soient les plus "frais" (et pertinents) possibles. Si on lance une requête sur un nom de sportif ou d'une compétition, on s'attend à afficher, comme servis sur un plateau, les derniers résultats de tel ou tel match ou compétition. C'est un test que l'on ne peut effectuer, pratiquement, qu'à partir des versions anglaises de ces 4 moteurs de recherche : Ask, Bing, Google et Yahoo! Prenez cet exemple de recherche puis voyez comment ils s'ingénient à nous présenter des résultats dédiés comportant des fonctionnalités plus ou moins avancées : "f1".

Related searches for f1
[McLaren F1](#) [F1 Cars](#) [ITV F1](#)
[F1 Ferrari](#) [F1 Babe](#) [Worst F1 Crashes](#)
[More related searches »](#)

[Formula One 2009](#)
 Formula One, abbreviated to **F1** and also known as Grand Prix racing, is the highest class of single-seat open-wheel formula auto racing. It... [More »](#)
[Source](#) Go To: [Calendar](#) · [Standings](#) · [Teams](#) · [History](#) · [Results](#)
Races:
Drivers:

Ask.com

Best match
[The Official Formula 1 Website](#)
www.formula1.com · Official site
 Formula One, abbreviated to **F1**, is the highest class of open wheeled auto racing defined by the Fédération Internationale de l'Automobile...

2009 Calendar	Season
News	Results
Live Timing	Headlines
Technical	Races

[Formula One racing](#) 2009

May 24	Monaco Grand Prix	Jenson Button	78 laps	Brawn #22
Jun 7	Turkish Grand Prix	Jenson Button	58 laps	Brawn #22
Jun 21	British Grand Prix	Sebastian Vettel	60 laps	Red Bull #15
Jul 12	German Grand Prix	8:00 AM ET, FOX	60 laps	191.94 mi
Jul 26	Hungarian Grand Prix	8:00 AM ET, FOX	70 laps	190.56 mi


[Schedule](#) · [FOX Sports on MSN](#) ·
 Statistics provided by STATS, Inc. © 2009

Bing

Google [Advanced Search](#)
[Preferences](#)

Web [Show options...](#) Results 1 - 10

[Formula 1 Racing: Season standings](#)

	Driver	Team	Points
	1. Jenson Button	Brawn	64
	2. Rubens Barrichello	Brawn	41
	3. Sebastian Vettel	Red Bull	39

Next race: German Grand Prix, Jul 12 8:00am
www.formula1.com


Google

Etant donné que Yahoo! ne propose pas de raccourcis de recherche pour cette requête, voici d'autres exemples de résultats dédiés au tennis (Google, Yahoo! et Ask)...

Google [Advanced Search](#)
[Preferences](#)

Web [Show options...](#) Results 1 - 10

Wimbledon Finals: Roger **Federer** def Andy Roddick

	Roger Federer (2)	5 7 7 3 16
	Andy Roddick (6)	7 6 6 6 14

Google

Also try: [roger federer](#), [roger federer girlfriend](#), [More...](#)

Wimbledon - Monday July 6, 2009
[Scores and Schedule](#) | [Photos](#) | [Video](#)

Men's Singles	Set 1	Set 2	Set 3	Set 4	Set 5
(2) Roger Federer	5	7 (8)	7 (7)	3	16
(6) Andy Roddick	7	6 (6)	6 (5)	6	14

[sports.yahoo.com](#) [More Results](#)

Roger Federer - [Wikipedia, the free encyclopedia](#)
[Childhood and...](#) | [Junior tennis](#) | [Early career...](#) | [Top 10 and...](#)


Roger **Federer** is a Swiss professional tennis player. He is currently ranked World No. 2 having held the World No. 1 position for a record 237 consecutive weeks. Most sports analysts, tennis critics, and former players...

[en.wikipedia.org/wiki/Roger_Federer](#) - 375k - [Cached](#)



Yahoo!


Tennis - [Wikipedia, the free encyclopedia](#)


 **Tennis** is a sport played between two players (singles) or between two teams of two players each (doubles). Each player uses a strung racquet to strike a hollow rubber ball covered with felt (most of t...

[Source](#) [en.wikipedia.org/wiki/Tennis](#)

TV listings for tennis

Zip Code: 10010 · **Provider:** Time Warner Southern Manhattan-Digital (New York) [Edit](#)

 Fri , 10:00 AM ET · [Tennis](#) : Davis Cup: Croatia vs. United States
 · [Channel 455](#) · [TENNIS](#) · Duration: 180 min.

 Fri , 1:00 PM ET · [Tennis](#) : Davis Cup: Croatia vs. United States
 · [Channel 455](#) · [TENNIS](#) · Duration: 180 min.

[Ask TV](#) · [TV data from Zap2it](#) · [More TV listings »](#)

Questions about tennis

[What is tennis player Safinams first name](#)
[What is tennis elbow?](#)
[How to Play Tennis](#)
[Ask Q&A](#) · [See all 23,636 questions about tennis »](#)

Ask.com

Nous avons ajouté Ask car c'est le seul moteur à afficher directement les heures des retransmissions télévisées. Au final, la plupart des moteurs s'ingénient, à travers la recherche universelle, à nous proposer des résultats en temps réel : actualités, articles de blogs, pages de forums, etc. Reste à savoir s'ils seront capables de structurer ces informations de façon à développer des raccourcis qui ressembleront presque à des sortes de mini-pages web (un peu à l'image de *Wolfram Alpha* ou de ce que propose *Google Squared*). Nous trouvons que Bing et Ask sont sur la bonne voie !

Les moteurs de recherche en temps réel

Voici maintenant des moteurs qui, le plus souvent, sont strictement anglophones et penchent nettement vers les réseaux sociaux...

Collecta (<http://collecta.com>) est un moteur qui scrute les résultats issus des blogs, des commentaires de blogs, des articles récents, des sites de micro-blogging (Twitter, Jaiku, Identica) et des images postées sur Flickr. Cliquez sur le lien *More options* afin d'opérer un tri rapide. Cliquez sur un des liens qui sont visibles afin d'afficher l'article correspondant dans le volet de droite. Vous pouvez à tout moment interrompre la collecte des données en cliquant sur l'icône *Pause*. Un lien vous permet de relancer une recherche dans Google. On peut s'étonner des problèmes de rendu de mise en page dans la plupart des navigateurs et du fait

que ce moteur ne soit pas compatible avec Internet Explorer. Ça fait beaucoup de problèmes à la fois !

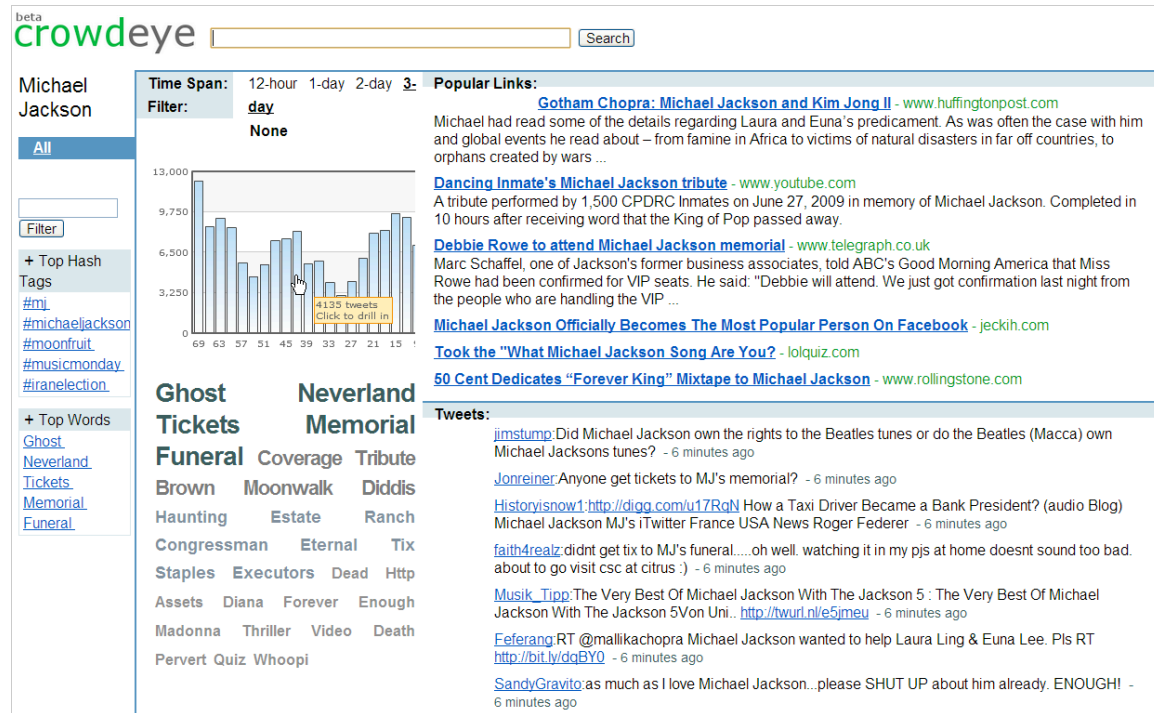
Crowdeye (<http://www.crowdeye.com/>) est un moteur qui nous a paru plus intéressant que le précédent. Il offre des possibilités de syntaxe avancée : [jackson OR #michaeljackson - \(Apple iPod\) OR \(Microsoft Zune\) - "LA Lakers" OR Lakers OR "Orlando Magic" OR "NBA Finals"](#).

Notez que les guillemets servent à lancer une recherche sur une expression exacte alors que les parenthèses vous permettent de regrouper des éléments d'une requête.

Le volet de gauche liste les *Hashtags* ainsi que les mots-clés les plus utilisés.

Dans la page des résultats, vous avez la liste des entités nommées qui ont été détectées.

Les liens détectés sont classés en deux catégories : les articles les plus populaires et les messages postés sur Twitter.



Vous pouvez classer les résultats en fonction de leur ancienneté. Le graphe qui est visible permet d'afficher directement les données recueillies par le moteur. Cliquez simplement sur le jour qui vous intéresse.

Impints (<http://www.itpints.com>) est un outil intéressant dans le sens où vous pouvez choisir le type de contenu à analyser en cliquant sur le lien *Advanced Search* et qu'il est possible de récupérer un flux RSS pour chacune des pages de résultats générées.

Spy (<http://spy.appspot.com>) est une application Google App Engine qui vous permet d'être à l'écoute des conversations dans les sites de réseautage social et de Lifestreaming : Twitter, FriendFeed, Flickr, commentaires laissés sur BackType, Yahoo News, les blogs et Google Reader. Il est possible de récupérer l'adresse du flux RSS pour chacune de vos recherches.

Keotag (<http://www.keotag.com>) est un méta moteur de recherche qui vous permet de choisir votre source d'informations : Technorati, delicious, Twitter et bien d'autre encore... Pour chacune des recherche, il est possible de récupérer la liste des liens au format OPML ou le fil d'informations correspondant.

OneRiot (<http://www.oneriot.com>) est un moteur qui se propose de traquer les dernières actualités, billets de blogs, vidéos et contenus issue du Web social (Twitter, Digg, etc.). Vous pouvez classer les résultats en utilisant ces deux filtres :

- *Realtime* : les liens affichés seront ceux qui ont été les plus récemment partagés ;
- *Pulse* : les liens sélectionnés présenteront, cette fois-ci, l'implication sociale la plus forte.

OneRiot beta

Search the realtime web

Web Video

google wave

Search

Trending Topics: Air France, Follow Friday, Economy, David Carradine, Barack Obama, World Environment

Most relevant results: **Friday, Jun 5 at 2:59pm** Sort by: Pulse | **Realtime**

Google Wave for making a lot money.
Google Wave for making a lot money. Waving For Business Google Wave has obvious business uses, combining tools that businesspeople use every day. It...
<http://www.mmogtraders.com/googlewave.htm>

Last shared: 3 minutes ago | 95 shares (expand) | Tweet It First shared on Twitter by Googleg209

Google Wave: First Impressions -- Google -- InformationWeek
Google Wave: First Impressions Wave's potential as open source collaborative plumbing for the Internet could help expand it beyond a simple product and...
<http://www.informationweek.com/news/software/development/showArticle.jhtml?articleID=2...>

Last shared: 2 hours 39 minutes ago | 26 shares (expand) | Tweet It First shared on Twitter by twittlord

Experience the World's first Google Wave SVG Editor Gadget - The Next ...
You are here: Home News Experience the World's first Google Wave SVG Editor Gadget Experience the World's first Google Wave SVG Editor...
<http://thenextweb.com/2009/06/05/experience-worlds-google-wave-svg-editor-gadget/>

Last shared: 34 minutes ago | 10 shares (expand) | Tweet It First shared on Twitter by 24z

Most shared today

4 travelers missed doomed Air France Flight 4...
Yahoo! News

First shared by NewsWorldNow

Obama Arab World Speech: Extremists Call Him ...
Huffington Post

First shared by mabs0

CARRADINE UPDATE: UNNATURAL CAUSES | FilmDrunk...
filmdrunk.uproxx.com

First shared by FilmDrunk

BBC NEWS | Business | Sir Alan hired in gover...

Parmi les facteurs qui sont déterminants figurent le nombre de fois où un lien a été partagé et l'indice de confiance des personnes qui ont partagé ces liens (ce n'est qu'une partie de l'algorithme).

Si aucun flux RSS n'est généré, il existe de nombreuses façons d'intégrer OneRiot aux outils que vous utilisez habituellement (<http://www.oneriot.com/product/addons>) :

- Gadgets à ajouter à des services de pages personnalisées : MyYahoo!, iGoogle et My MSN ;
- Barre d'outils Facebook ;
- Intégration des résultats de OneRiot dans Google grâce à une extension appelée WebMynd ;
- Intégration de OneRiot à une autre extension appelée SurfCanyon ;
- Intégration d'une barre d'outils dédiée à voter navigateur : PulseChecker ;
- Intégration de OneRiot à MySpace ;
- Utilisation des Web Slices et des suggestions de recherche avec Internet Explorer 8.

Socialmention (<http://socialmention.com>) est un excellent moteur qui propose un grand nombre de filtres (date, source), des possibilités d'export élaborées (RSS, alerte e-mail, fichier CSV), un module de recherche avancée ainsi que des statistiques relativement complètes (tonalité des articles, mots-clés employés, liste des internautes, Hashtags utilisés, classification des sources).

socialmention*

Blogs Microblogs Bookmarks Comments Events Images News Video Audio Q&A All

android

Search

[Advanced Search](#)
[Preferences](#)

39% strength

2:1 sentiment

20% passion

6% reach

19 minutes avg. per mention

last mention 29 minutes ago

8 unique authors

Mentions about android

Sort By: Date Results: Anytime Results 1 - 15 of 80 mentions.

BenQ plans to launch Android-powered smartphone and netbook next year

IDG News is reporting that Taiwanese consumer electronics manufacturer BenQ announced Thursday its intentions to launch both an Android-based smartphone and a netboo...
www.mobileburn.com/news.jsp?id=7201
29 minutes ago - on [yahoo](#)

Links to GreyBlogs.com

greyblogs.wordpress.com/links-to-greyblogs-com/
46 minutes ago - by greyblogs on [wordpress](#)

Paranoid Android

annebgodbout.wordpress.com/2009/06/05/paranoid-android/
58 minutes ago - by annebgodbout on [wordpress](#)

Berita : Windows 7 Final Bakal akan Dirilis!

sl4yit.wordpress.com/2009/06/05/berita-windows-7-final-bakal-akan-dirilis/
60 minutes ago - by sl4y3r1995 on [wordpress](#)

Acer's Android Netbook Will Come With Windows, Fail at Being an "Android Netbook" [Android] (Gizmodo)

Sentiment

positive

neutral

negative

5

72

3

Top Keywords

netbook

windows

subscribe

explore

gadgets

google

benq

source

company

smartphone

17

7

6

5

5

5

5

4

4

4

Voici un exemple de syntaxe :

[http://socialmention.com/search?q="Susan+Boyle"&t=blogs&f=rss](http://socialmention.com/search?q='Susan+Boyle'&t=blogs&f=rss). Il est possible d'afficher les tendances (<http://socialmention.com/trends>) ainsi que d'utiliser une API et générer des fichiers au format JSON ou PHP (<http://code.google.com/p/socialmention-api/wiki/APIDocumentation>). Enfin, vous pouvez ajouter Socialmention à votre liste de moteurs personnalisés en cliquant sur le lien *Search plugin*. C'est réellement la bonne surprise de cette sélection...

Twingly (<http://www.twingly.com>) est à la fois un moteur de recherche dans les blogs et les micro-blogs. Cliquez tout d'abord sur les liens *Twingly Blog Search* et *Advanced search*. Trois options sont intéressantes :

- *Return posts within the blog (blog:)* : affiche toutes les pages qui font un lien vers un blog en particulier. Il vous suffit ensuite de cliquer sur le lien *n inlinks*.
- *Return posts that link to the page or site (link:)* : affiche tous les messages qui font un lien vers une page d'un blog ou d'un site en particulier.

The screenshot shows the TwinglyRank search interface. At the top, there's a search bar with the query 'link:http://abondance.com'. To the left is the TwinglyRank logo. Below the search bar, it says 'Hot right now: Cairo University David Carradine Egypt Tiananmen Square'. The main section is titled 'Search results for "link:http://abondance.com"'. It includes filters for 'Sort by: TwinglyRank', 'Results from: Any source', 'Anytime', 'All languages', and a 'Reset filters' button. Three results are listed:

- 1** Links/likes: [Un tableau de bord pour le Local Business Center de Google](#) (2 days and 4 hours ago in [Abondance](#), [Actualités Google](#)). Description: Le Local Business Center de Google, qui permet de gérer l'affichage des données professionnelles dans Google Maps, propose maintenant un tableau de bord pour suivre les statistiques d'affichage de c...
Profile: [Secrets2Moteurs : Veille Référencement & Moteurs 1.inlink](#)
- 0** Links/likes: [Google Squared veut des résultats bien carrés...](#) (1 day and 9 hours ago). Description: Google Squared a pour vocation de classer les résultats de recherche dans un tableau comparatif, proposant des données chiffrées ou des informations structurées par rapport à la requête demandée... L'ou...
Profile: [Le Blogueur](#)
- 0** Links/likes: [Google ne tiendrait plus compte de l'attribut Nofollow pour la tr...](#) (1 day and 8 hours ago). Description: Selon Matt Cutts, porte-parole SEO de Google, Google ne prendrait plus en compte la technique du PageRank Sculpting au travers de l'attribut "nofollow" des liens, qui était devenue une technique de ré...
Profile: [Le Blogueur](#)

- *Return posts from the site or domain (site:)* : affiche tous les messages d'un blog en particulier.

Vous pouvez ensuite les classer par date, en fonction de l'importance des sources (*TwinglyRank*), exclure les blogs contenant du Spam, classer les sources par date ou en fonction de la langue d'interface. Enfin, il est possible de souscrire une alerte e-mail ou au flux RSS qui est, à chaque fois, généré.

Twingly Microblog permet d'effectuer une recherche dans les sites de micro-blogging comme Twitter, Jaiku, Identica, etc. Ces opérateurs sont autorisés :

- *From:* : les messages postés par un utilisateur ;
- *To:* : les messages envoyés à un utilisateur ;
- *Since:* : les messages postés de telle à telle date ;
- *Until:* : les messages postés jusqu'à cette date ;
- *#* : les messages marqués d'un HashTag en particulier ;
- *@* : les messages qui citent un utilisateur en particulier.

L'intérêt par rapport au module de recherche proposé par Twitter n'est pas évident tant ce service domine les autres. Pour vous en convaincre, décochez la case Twitter puis cliquez sur le bouton *Update results*.

TipTop (<http://www.feeltiptop.com>) se définit comme une plateforme d'analyse sémantique en langage naturel et appliquée à Twitter. Après avoir lancé une recherche (*Michael Jackson*, par exemple), vous allez constater que vous pouvez filtrer les résultats de beaucoup de façons :

- En utilisant les mots-clés affichés dans le volet de gauche (*memorial, ticket, music, death*, etc.) ;

Le moteur va alors procéder par adjonction de termes.

- En cliquant sur les onglets visibles au-dessus des résultats : citations de lieux, de produits, de personnes ou de liens vers des pages web.

La pertinence des filtres reste discutable mais ça n'a rien d'étonnant...

The screenshot shows a Twitter search results page for the query "michael jackson". The interface includes a filter bar at the top with tabs for "Places", "Types of products", "Types of events", and "Web pages". Below this, there are radio buttons for "Top Tips and Pits" (selected) and "Remaining messages". The results are divided into two columns: "Tips (17%)" and "Pits (14%)".

Left Column (Tips):

- Tip 1:** "Holy shit!! I got Michael Jackson memorial service tix at staples! Aaahhhh!! man in mirror smiling!!" with a link to <http://yfrog.com/7goizi>, posted 3 minutes ago.
- Tip 2:** "I love to live in a world where Entertainment Weekly does an entire, quality issue dedicated to Michael Jackson." posted 1 minute ago.
- Tip 3:** "FOR 10 FREE Tickets To Michael Jackson Memorial. Check out www.rulabrown.com" posted 2 minutes ago.

Right Column (Pits):

- Pit 1:** "Michael Jackson's messy finances provide some harsh lessons. \$4.5 mill monthly interest payments:" with a link to <http://twurl.nl/bk5x0d>, posted 3 minutes ago.
- Pit 2:** "www.aroundtheway.com has footage of michael jackson's ghost" with a link to <http://www.aroundtheway.com/videos/view/203>, posted 1 minute ago.
- Pit 3:** "watching the TMZ live stream of Michael Jackson's court hearing.. so sad people are fighting who gets what..." posted 2 minutes ago.

On the left side of the page, there is a sidebar with a "Filter results by" dropdown (set to "No Filter"), a list of tags (e.g., "jackson", "michael", "memorial", "ticket", "ghost", "service", "video", "funeral", "news", "people", "staples", "music", "world", "tv", "youtube", "work", "court", "circus", "adventurous", "beach", "italy", "venture", "entrant", "man", "ottawa"), and an "Ads by Google" section for "King of POP M J DVD \$74.95".

La page des résultats est divisée en deux colonnes :

- Les tips : regroupent les messages à connotation positive qui offrent des recommandations : tuyaux, bons plans, choses à faire, etc.
- Les pits : sont des contributions plutôt négatives et qui indiquent généralement quelles sont les choses à ne pas faire.

Les barres vertes et rouges indiquent le pourcentage des tips ou des pits comparés à l'ensemble des tweets détectés.

Activez le bouton radio *Remaining messages* afin d'afficher les messages non classés. On peut tout de même penser que les équipes de TipTop ont encore beaucoup à faire avant que le moteur soit à la hauteur des attentes suscitées par leur "pitch". Et nous ne parlons pas des publicités AdSense qui n'aident pas à la compréhension des résultats.

Signalons également ces autres outils de recherche : **Scooper** (<http://www.scooper.com>), **Yauba** (<http://fr.yauba.com>), **WhosTalkin** (<http://www.whostalkin.com>), **Samepoint** (<http://www.samepoint.com>) ou encore **IceRocket** (<http://www.icerocket.com/search?tab=twitter>).

Daillyrt (<http://www.dailyrt.com>) est un portail dédié à Twitter (dont nous n'avons pas parlé dans la première partie de cet article) et qui offre un classement des tweets en fonction du nombre de fois qu'ils ont été retweetés. Deux flux RSS sont proposés (TOP 10 et TOP 20 des tweets les plus populaires). Par ailleurs, vous pouvez classer les tweets en fonction du nombre de Followers que compte leur auteur et de leur type de contenu (vidéos, images, liens).

The screenshot shows the homepage of the "dailyrt" website, which is described as "The most popular tweets on the web." The interface includes a navigation bar with tabs for "hot tweets", "live tweets", "resources", and "trends". Below this, there is a search bar with the query "michael jackson" and a "Search" button. The results are displayed in a list format, showing the tweet text, the user's profile picture, the number of retweets, and the time of posting.

Search Results:

- 1st result:** "Michael van Poppel mpoppel Jul 6, 2009, 12:08 pm Michael Jackson's 'GHOST' seen on CNN during live shot: <http://bit.ly/WT4CR>" with 7 retweets.
- 2nd result:** "El Online online Jul 6, 2009, 12:00 pm Judge Removes Mom From Control of Michael Jackson's Estate <http://tinyurl.com/nsfh38>" with 5 retweets.
- 3rd result:** "Kristen Nicole KristenNicole2 Jul 6, 2009, 11:59 am from allfacebook: Michael Jackson Officially Becomes The Most Popular Person On Facebook: This .. <http://tinyurl.com/lqnezo>" with 2 retweets.

On the right side of the page, there is a "Filter Your Search..." section with checkboxes for "Containing:" (Links, Videos, Images), a "Follower Limit:" section with "Max Followers:" and "Min Followers:" input fields, and a "Dates:" section with "Since this date:" and "Until this date:" input fields.

Cette liste n'est sûrement pas exhaustive !

Créer votre propre moteur de recherche en temps réel

Il est très facile d'utiliser un moteur personnalisé Google CSE (<http://www.google.fr/coop/cse/>) afin de limiter les résultats de vos recherches aux dernières 24 heures. Voici un exemple de code possible :

```
<form action=" http://www.exemple.com/rechercher.html" id="cse-search-box">
  <div>
    <input type="hidden" name="cx" value="016121005022341368769:bifbfdj87xi" />
    <input type="hidden" name="cof" value="FORID:10" />
    <input type="hidden" name="ie" value="UTF-8" />
    <input type="text" name="q" value="android" size="31" />
    <input type="hidden" name="as_qdr" value="d" />
    <input type="hidden" name="num" value="100" />
    <input type="submit" name="sa" value="Rechercher" />
  </div>
</form>
<script type="text/javascript" src="http://www.google.com/coop/cse/brand?form=cse-
search-box&lang=fr"></script>
<div id="cse-search-results"></div>
<script type="text/javascript">
  var googleSearchIframeName = "cse-search-results";
  var googleSearchFormName = "cse-search-box";
  var googleSearchFrameWidth = 600;
  var googleSearchDomain = "www.google.com";
  var googleSearchPath = "/cse";
</script>
<script type="text/javascript"
src="http://www.google.com/afsonline/show_afs_search.js"></script>
```

On s'est contenté de passer ces paramètres supplémentaires :

- `<input type="text" name="q" value="android" size="31" />` : ajoute un mot-clé par défaut qui s'affichera dans la barre de recherches ;
- `<input type="hidden" name="as_qdr" value="d" />` : ajoute un autre commutateur qui forcera l'affichage des résultats indexés durant les dernières 24 heures ;
- `<input type="hidden" name="num" value="100" />` : permet d'afficher 100 résultats par page.

En pratique, rien n'est simple puisque l'API Google Search ne permet pas de restreindre les résultats par date et que vous ne pourrez pas utiliser l'API Google Ajax pour créer un moteur en Ajax (le filtre par date ne fonctionnera pas).

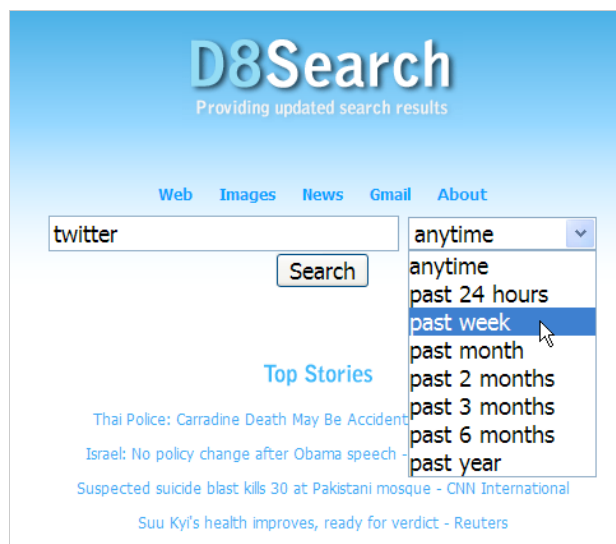
L'intérêt est donc plus de permettre une recherche dans une liste de sites que vous avez déterminés et de renforcer la pertinence des résultats en apposant différents filtres par date.

Il existe une bonne démonstration de ce type de procédé à cette adresse :

<http://d8search.com>.

Voici un exemple du code utilisé :

```
<input type="text" name="q" size="30" />
&nbsp;<select name="as_qdr">
  <option value="all">N'importe quelle date
  <option value="d">Dernières 24 heures
  <option value="w">Semaine passée
  <option value="m">Mois passé
  <option value="m2">Les 2 derniers mois
  <option value="m3">Les 3 derniers mois
  <option value="m6">Les 6 derniers mois
  <option value="y">L'année dernière</select>
```



Archiver les pages générées par une recherche

C'est un peu la question que nous nous sommes posée après avoir créé un moteur de recherche personnalisé. Non pas forcément pour sauvegarder des pages d'un moteur CSE mais plutôt afin de générer automatiquement une archive après avoir saisi une requête en particulier. Le principe consiste, par exemple, à utiliser une librairie PHP appelé cURL pour automatiser cette procédure.

cURL ("*Client URL Request Library*") permet de récupérer le contenu d'une ressource hébergée sur un serveur web. cURL supporte notamment les protocoles FTP, FTPS, HTTP, HTTPS, TFTP, SCP, SFTP, Telnet, DICT, FILE et LDAP. L'écriture peut se faire en HTTP en utilisant les commandes POST ou PUT.

Pour le test en local que nous avons effectué, nous nous sommes servis de ce serveur web : **XAMPP** (<http://www.apachefriends.org/fr/xampp-windows.html>).

Une fois que vous avez installé XAMPP Lite, vous devez activer cURL en éditant le fichier *PHP.ini* qui se trouve dans `\xampplite\php`. Il suffit ensuite de supprimer le point-virgule qui précède cette commande : `extension=php_curl.dll`. Notez que quelque soit le serveur que vous utilisez (en local ou à distance), la procédure est, à peu de chose près, la même.

1. Créez ensuite un fichier PHP dans `\xampplite\htdocs\site01`.
2. Collez ce contenu :

```
<html>
<body>
<?php
$today = date("dmY");
set_time_limit(0);
ini_set('display_errors',true);
$fp = fopen (dirname(__FILE__) . '/test'.$today.'.html', 'w+');
$ch =
curl_init('http://www.google.com/search?q="olivier+andrieu"&hl=en&lr=lang_fr&tbo=1&num=100&tbs=qdr:d,sbd:1');
curl_setopt($ch, CURLOPT_TIMEOUT, 50);
curl_setopt($ch, CURLOPT_FILE, $fp);
curl_setopt($ch, CURLOPT_FOLLOWLOCATION, true);
curl_exec($ch);
curl_close($ch);
fclose($fp);
?>
```

Nous lançons une recherche dans Google qui va nous permettre de récupérer les résultats en français sur la requête "Olivier Andrieu" et en filtrant les résultats aux dernières 24 heures.

Nous nous sommes librement inspirés d'un script qui est visible à cette adresse :
<http://www.webdigity.com/index.php?action=tutorial;code=45>.

Après avoir démarré les modules Apache et MySQL, ouvrez votre navigateur puis saisissez l'adresse URL en local du fichier : <http://localhost/test.php>.
La page des résultats ([Test+date_du_jour.html](#)) sera automatiquement enregistrée à l'emplacement à partir duquel vous avez lancé votre script.

- `$today = date("dmY");` : permet de définir un format de date et de créer une variable nommée \$today ;
- `set_time_limit(0);` : évite les problèmes d'expiration de délai lors du téléchargement de fichiers volumineux ;
- `ini_set('display_errors',true);` : signale la cause d'une erreur possible ;
- `fopen(dirname(__FILE__))` : ouvre le fichier ou l'adresse URL qui sera indiquée ;
- `"/test".$today.'.html'` : génère un fichier dans lequel la date du jour sera mentionnée ;
- `'w+'` : précise le mode d'ouverture du fichier (la valeur est w pour "Écriture") ;
- `curl_init` : démarre une nouvelle session et retourne un identifiant de session (ch) ;
- `curl_setopt` : définit les options de transfert de la session cURL ;
- `CURLOPT_TIMEOUT` : indique le temps maximal de la session cURL ;
- `CURLOPT_FILE` : fonctionne comme un descripteur de fichier dans lequel seront envoyées les données ;
- `CURLOPT_FOLLOWLOCATION` : reçoit la valeur TRUE afin de suivre les redirections http ;
- `curl_exec` : exécute la session cURL définie par ch ;
- `curl_close` : ferme la session cURL.

Voici un exemple de page une fois qu'elle a été enregistrée :

The screenshot shows a Google search interface. At the top, the search bar contains "olivier andrieu" and the "Search" button is visible. Below the search bar, there are radio buttons for "Search the Web" and "Search French pages", with "Search French pages" being selected. The search results are displayed below, showing a list of links and snippets. The first result is "Amomenti » Blog Archive » Twitter or not ? :: Stratégies web ...". The second result is "Livres Informatique d'entreprise - Librairie Eyrolles". The third result is "Résultats ApocalX Mix pour Comment conserver son référencement ...". The fourth result is "L'association des professionnels de l'information et de la ...". On the right side of the search results, there is a section for "Sponsored Links" featuring "Olivier Andrieu" with a link to "Amazon.fr/livres".

Bien entendu, les possibilités offertes par cURL sont quasi infinies (notamment à l'aide des requêtes XPath) et un détour sur le site officiel et la documentation en ligne est obligatoire : <http://curl.haxx.se/>.

Signalons pour terminer qu'il est possible de faire beaucoup de choses avec du PHP comme, par exemple, extraire automatiquement les URLs des résultats de recherche dans Google : <http://www.seoblackout.com/2008/10/26/extraire-resultats-google/>. Bien que la méthode utilisée fasse appel à d'autres fonctions, la démarche d'esprit reste la même.

Le règne de l'information volatile ?

Nous ne croyons pas vraiment au journalisme de type "micro-blogging" qui vous permet de suivre en temps réel une catastrophe aérienne ou les prémices d'une révolution. Il est sûr que vous pouvez, grâce à Twitter, prendre le pouls d'un événement mais vous aurez forcément

besoin, à un moment ou à un autre, de prendre de la hauteur par rapport au flux continu des nouvelles. De fait, un monde idéal consisterait à offrir, dans un mouvement simultané, les points de vue à chaud et les analyses de fond sur une actualité donnée.

D'un côté, nous avons Google qui offre des possibilités de recherche dans le temps et nous permet, à partir de la version anglaise, de les trier par date, mais qui ne génère aucun flux RSS, bride la combinaison de cette fonctionnalité avec l'emploi de certains opérateurs et ne l'a pas encore intégrée à son API. Et, de l'autre, Twitter qui offre des possibilités de syntaxe avancées ainsi qu'un affichage en temps réel des résultats mais sans, pour le moment, les classer par pertinence (un peu à l'image de que fait Digg). L'avantage de Google est qu'il s'appuie sur des algorithmes puissants et scanne une grande partie du Web tandis que Twitter fonctionne à la manière d'un réseau social dans lequel chacun est mis à contribution et apporte sa pierre à l'édifice. Puisque ce sont les utilisateurs de Twitter qui font office de lien entre les internautes et les pages web, on peut donc logiquement penser que leurs contributions seront aussi évaluées en fonction d'un critère - que l'on appellera comme on le voudra -, mais qui permettra de faire la part entre les utilisateurs "de confiance" et les autres.

Au final, la recherche en temps réel offre les avantages suivants :

- Dans certains cas, une plus grande pertinence des résultats quand une requête est directement liée à une actualité récente : "ml" pour tout savoir des circonstances de la mort de Michael Jackson et non afin d'accéder au site officiel ;

- Des pages nouvelles par rapport à un sujet que l'on connaît déjà dans ses grandes lignes : nous cherchons des pages citant des ressources d'icônes gratuites mais qui ne sont pas les mêmes que celles que nous avons déjà sauvegardées ou l'avis de tel ou tel blogueur plutôt que les documents classiques sur ce même sujet. On est très proche d'un processus de veille puisque le propos est moins de tout connaître d'un sujet que d'être au courant des nouveaux articles qui ont été écrites sur ce sujet.

- Une gestion de sa réputation numérique : nous sommes à la tête d'un hôtel réputé du bord de mer et souhaitons tout connaître des avis et des commentaires de nos clients ("*savoir ce qui s'est dit de nous*").

On peut s'étonner alors que Google ne propose pas ce type d'application "clé en main" alors qu'il possède tous les outils pour le faire : flux RSS, indexation en temps réel, outils de statistiques et de visualisation, etc. Pourquoi donc ne pas offrir une "OneBox" dédiée aux sites de réseautage social et de micro-blogging et qui nous donnerait la température et les tendances du moment sur telle ou telle thématique ? Peut-être n'est-ce qu'une question de mois, voire de semaines...

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/2009/07/la-recherche-temps-reel-avec-twitter.html>