

Améliorer ses recherches grâce aux moteurs verticaux de Google, Yahoo! et Microsoft (et les autres...)

[Retour au sommaire de la lettre](#)

Depuis quelques mois, sous l'impulsion de start-ups spécialisées comme Rollyo ou Eureka, les trois grands de la recherche d'information sur le Web ont créé des moteurs de recherche personnalisables nommés Google Custom Search Engine, Yahoo! Builder et Live Search Macros. Ces outils, facilement paramétrables, permettent en quelques minutes de mettre en ligne, pour une utilisation en solo, sur un intranet ou un site web public, un moteur de recherche vertical centré sur une thématique ou effectuant ses recherches sur un nombre limité et paramétrable de sites. Un gain de temps appréciable dans un processus de recherche ou de veille. Description des offres...

Depuis quelques mois, les grands acteurs de la recherche d'information que sont Google, Yahoo! et Microsoft ont mis en place une offre de recherche verticale, donnant la possibilité de créer, en quelques minutes, un moteur "sur mesure", effectuant ses recherches sur un certain nombre de sites web ou sur une thématique donnée.

Ces outils peuvent, bien sûr, être utilisés dans le cadre de sites web publics (voir par exemple des outils comme Biomalin (<http://www.biomalin.com/>) ou Grifil (<http://www.grifil.net/>), moteurs spécialisés sur le monde du bio ou de la BD, mais il en existe des centaines aujourd'hui...), mais également dans le cadre de projets plus personnels, à court terme (nécessité de mener à bien une étude sur une thématique précise pendant quelques jours ou quelques semaines) ou à plus long terme (veille dans un domaine précis)...

Il est effectivement très simple de se bâtir en quelques minutes, grâce à ces outils, son propre système de veille, permettant d'effectuer des requêtes dans une thématique très spécifique, voire sur un petit nombre de sites web de référence sur une problématique donnée... Voici comment faire avec les principaux outils disponibles sur le marché :

Google Custom Search Engine (<http://google.com/coop/cse/>)

L'outil de Google (qui nécessite d'avoir un compte - gratuit - chez le moteur de recherche) est d'une utilisation extrêmement simple : sur la page d'accueil, vous cliquez sur le bouton "Create a search engine") et vous définissez en quelques minutes la "customisation" de votre outil : son nom, sa description, des mots clés de recherche (qui permettront à Google d'affiner les requêtes), la langue utilisée, etc.

The screenshot shows the 'Create a Custom Search Engine' page on Google. The page is titled 'Google Custom Search Engine' and has a navigation menu on the left with links for 'Overview', 'New search engine', 'My search engines', 'My profile', 'Featured examples', 'Documentation', 'Google Marker', 'Custom Search Blog', 'Discussion group', and 'Support'. The main content area is divided into two steps: '1. Specify your search engine settings' and '2. Try it out and customize it further'. Under step 1, there are sections for 'Basic information' and 'Sites to search'. The 'Basic information' section includes fields for 'Search engine name' (with an example 'Real Climate Search Engine'), 'Search engine description' (with an example 'Climate Science from Climate Scientists'), and 'Search engine keywords' (with an example 'climate "global warming" "greenhouse gases"'). There is also a dropdown menu for 'Search engine language' set to 'English'. The 'Sites to search' section includes a text area for listing URLs, with a note 'List one URL per line'.

Il est ensuite demandé une liste de sites web sur lesquels porteront les recherches futures (vous pourrez en rajouter d'autres par la suite), avec deux choix :

- Soit les résultats futurs seront uniquement issus de ces sites web.
- Soit les résultats seront issus de l'index global de Google, mais avec une préférence pour les pages issues des sites indiqués.

Ainsi, selon vos besoins, vous pourrez opter pour l'une ou l'autre des solutions... Cependant, si vous optez pour une recherche générale (non limitée à la liste de sites fournie), nous vous recommandons d'ajouter sur cette page une liste de mots clés ayant rapport avec la thématique désirée dans le champ "Search engine keywords" afin d'affiner au mieux les résultats... Enfin, il est également possible de demander - ou non - l'affichage de publicités sur les pages de résultats. Pour les associations, agences gouvernementales ou universités uniquement... :-)

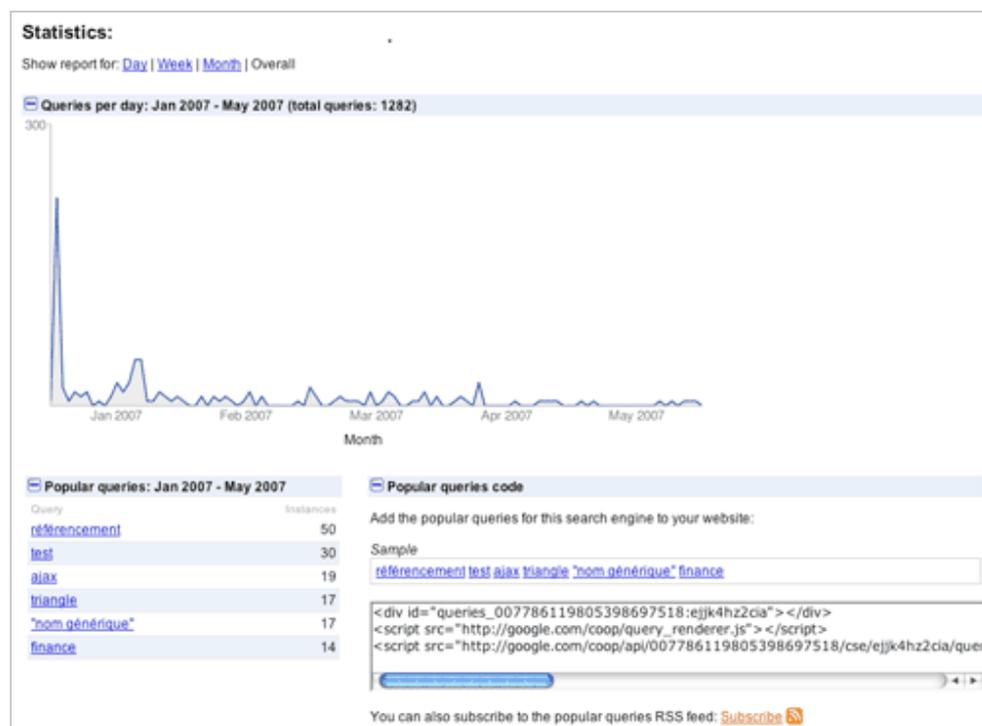
Autre solution très intéressante : la création de filtres de recherche sur l'url. Par exemple "*.gouv.fr" créera automatiquement un filtre sur tous les sites gouvernementaux français. Très utile... De nombreuses syntaxes sont proposées et explicitées sur le site.

Mais rien ne vous empêche également, à ce stade, de saisir uniquement l'adresse de VOTRE site web. Vous aurez ainsi en quelques secondes un moteur interne de bonne qualité avec la technologie de Google pour effectuer vos requêtes sur vos pages. Simple restriction : les recherches seront limitées au nombre de pages de votre site connues (et donc indexées) par Google, ce qui peut ne pas être exhaustif... Ce sera bien entendu le cas de tous les outils présentés dans cet article...

Après avoir testé que tout fonctionne bien grâce à un test de recherche, votre moteur est prêt à fonctionner. Durée de la création : moins de 5 minutes ! En fait, le plus long sera certainement de définir les mots clés et les sites sur lesquels porteront les recherches.

Une interface d'administration vous permettra ensuite de récupérer le code HTML du formulaire pour l'intégrer à votre intranet ou à une page web "perso", de modifier la charte graphique de l'outil, d'ajouter des thématiques ("labels") de recherche, etc. Bref, une batterie assez complète d'outils de création et de management de votre moteur personnalisé.

Vous aurez également à votre disposition un ensemble de statistiques simples, peu intéressantes pour une utilisation "en solo" du moteur mais qui peut avoir son importance dans le cadre d'un intranet par exemple :



L'avantage principal du système *Google Custom Search Engine* est son extrême simplicité. En quelques secondes, vous pouvez créer un moteur vertical sur un ou plusieurs site(s) web, sur une thématique donnée ou une problématique précise et bénéficier d'un moteur utilisant la technologie de Google dont la pertinence n'est plus à démontrer... Et cela s'avère très utile pour de nombreux projets portant, par exemple, sur un ensemble de quelques gros sites web de référence...

Yahoo! Search Builder (<http://builder.search.yahoo.com/>)

Yahoo! propose également un outil assez similaire à celui de Google, également accessible à tout titulaire d'un compte (gratuit) chez le portail.

L'outil proposé par Yahoo! propose 3 choix pour bâtir son moteur personnalisé :

- Recherche globale ou affinée sur le Web.
- Recherche sur un site donné.
- Recherche sur l'actualité (ce que ne propose pas Google).

YAHOO! SEARCH BUILDER [Search Builder Home](#) - [Help](#)

1 Define Search Engine 2 Design Search Box 3 Brand Search Results 4 Add to your site

Define search engine

Build a search engine that your customers can't find anywhere else. Allow your users to search the Web, your Site, or the News. Further enhance the experience by customizing the search results to fit your audience's needs.

Search Name
Name this search box. This is only for your reference; your users won't see this name.

* required

Search Types
Choose the types of search you want to allow on your site.

Search the Web [\[Customize the experience for your users\]](#)

Search this Site Domain:

Search Yahoo! News [\[Customize the experience for your users\]](#)

Default Search Type
Select which search type will be the default for your search box

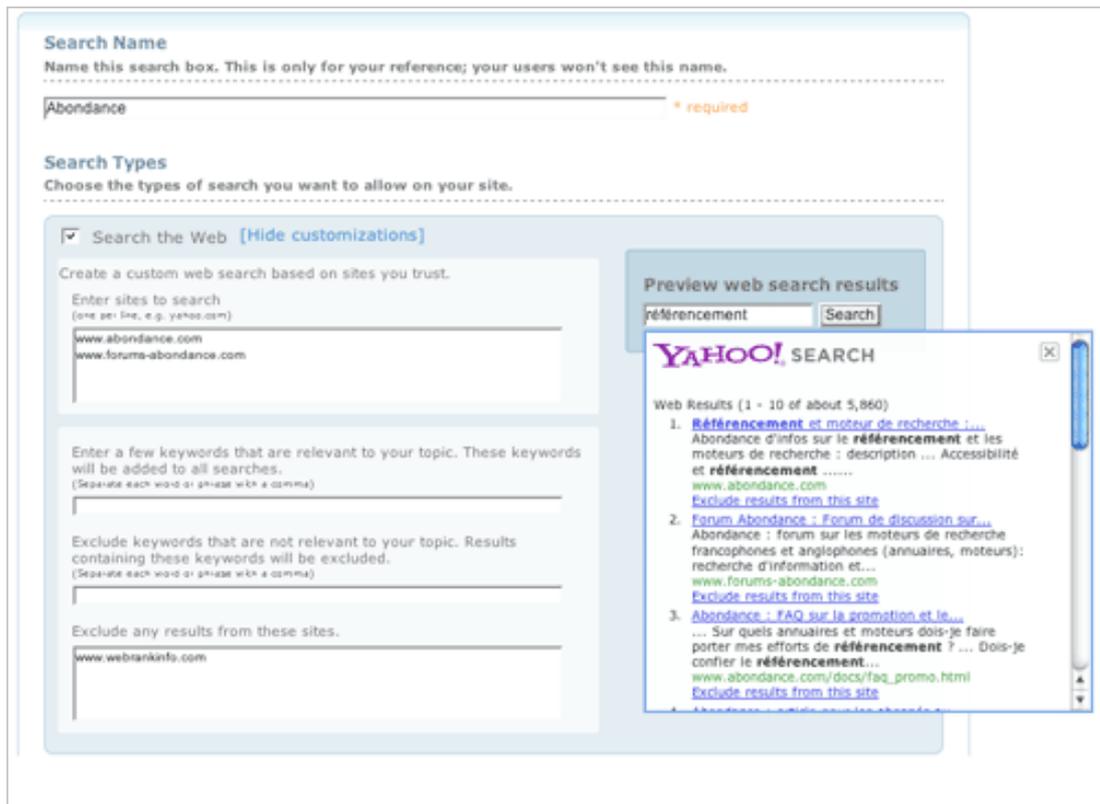
[Next](#) [Cancel](#)

Il est bien sûr possible de modifier le "look and feel" (couleur, taille, boutons radios, etc.) du formulaire, ajouter un logo, définir une langue, etc.

Mais le plus intéressant (et pourtant pas le plus visible, malheureusement) est le lien "*Customize the experience for your users*" qui permet :

- D'indiquer les sites web sur lesquels on désire effectuer les recherches.
- D'ajouter des mots clés de recherche.
- D'ajouter des mots clés d'exclusion (équivalent de la syntaxe "SAUF" qui sera ainsi ajoutée par défaut à toutes vos recherches, à manipuler avec précaution donc...)
- De créer une liste noire de sites web que vous ne voulez pas voir s'afficher dans vos résultats.

Une interface extrêmement pratique, très "Web 2.0" vous permet, au fur et à mesure de la création de votre outil, de vérifier sur le moteur quels résultats sont fournis :



L'outil est très bien fait et propose de nombreuses possibilités d'administration. Des statistiques sur les requêtes effectuées sur le moteur personnalisées sont également disponibles. Bref, un système proche de celui de Google au niveau des fonctionnalités mais plus étoffé de façon globale...

[Live Search macros \(http://search.live.com/macros/default.aspx?FORM=BJRE\)](http://search.live.com/macros/default.aspx?FORM=BJRE)

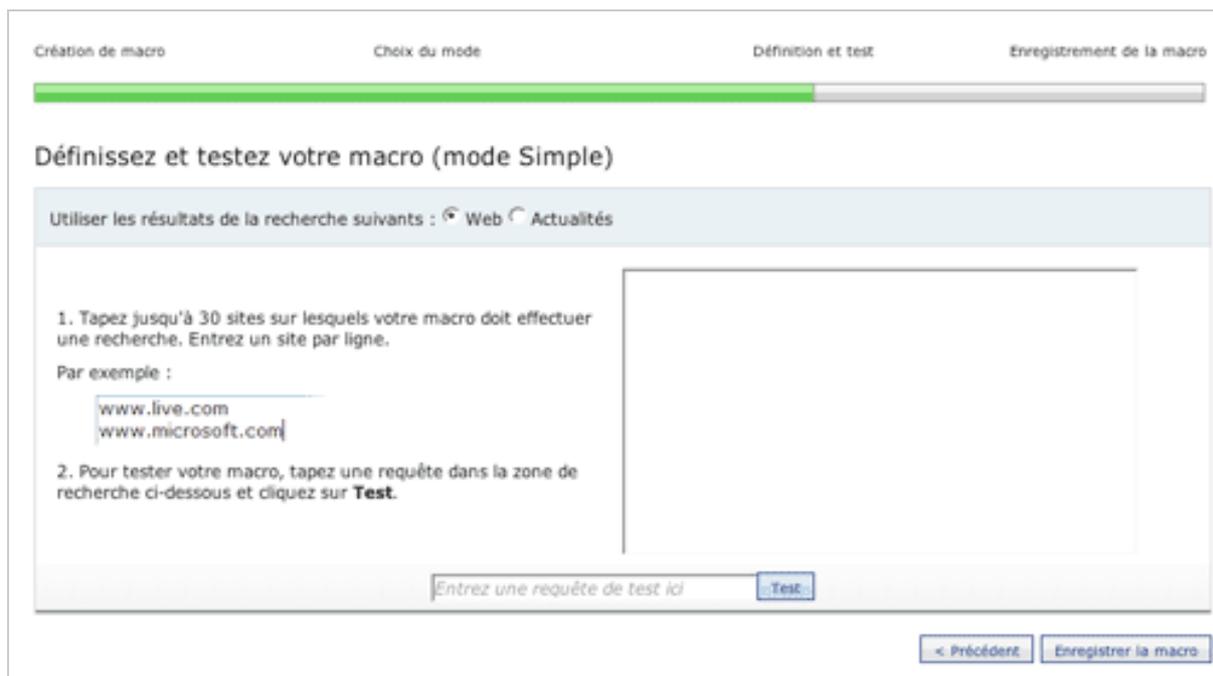
Live Search, le moteur de recherche de Microsoft n'est pas en reste avec ses "macros" (toujours disponibles aux titulaires d'un compte Hotmail, Messenger ou Passport), accessibles depuis l'onglet "Plus" puis "Modifier des macros" sur les pages de résultats :



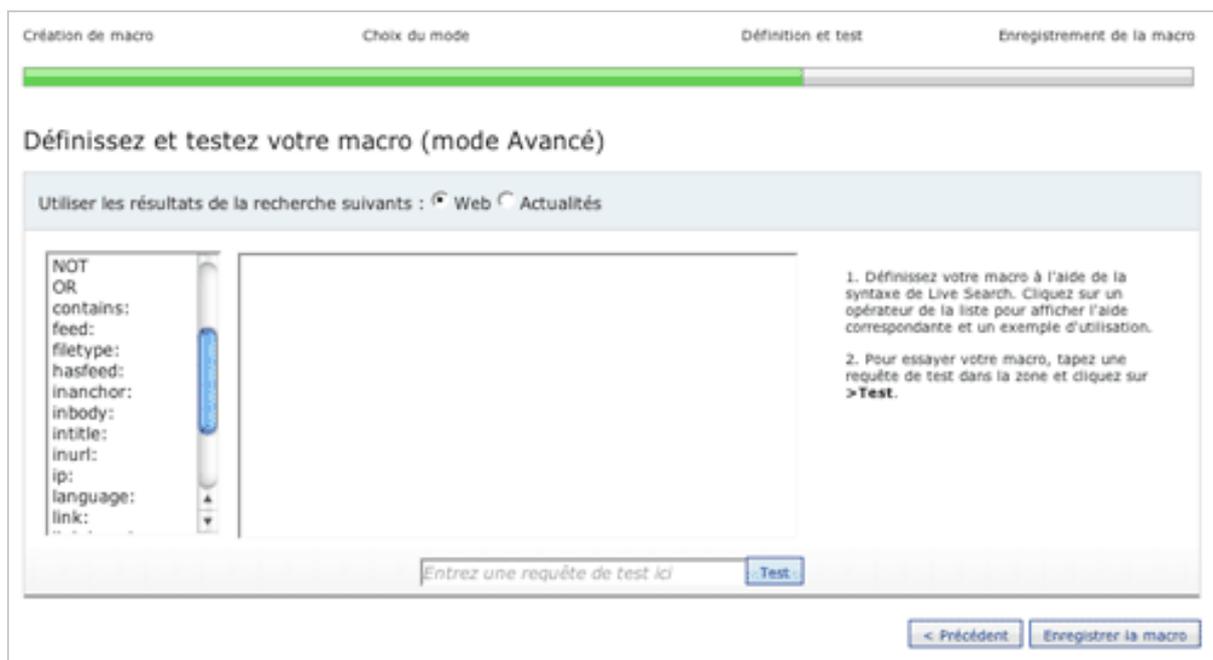
Deux choix sont ensuite proposés :

- Simple, où vous n'avez qu'à fournir une liste de sites sur lesquels porteront les recherches futures.
- Avancé, plus personnalisable.

L'interface "simple" est très... simple :-) sans être simpliste. Très claire, rapide, elle permet d'ajouter jusqu'à 30 sites web pour créer un moteur personnalisé.



L'interface avancée permet d'aller beaucoup plus loin en utilisant des syntaxes de recherche avancée (opérateurs booléens) et des opérateurs comme inurl:, feed:, link: ou autre.



On peut donc aller beaucoup plus loin qu'avec les autres outils en termes de personnalisation. Par exemple, effectuer des requêtes sur les pages qui contiennent un lien vers un document Word, ou celles qui contiennent un mot donné dans le titre, etc.

Seul problème avec cet outil : dès que nous avons créé une macro, quelle qu'elle soit, un message d'erreur ("Un problème est survenu lors du traitement de votre requête. Cliquez sur le bouton Précédent et réessayez. Si le problème persiste, réessayez ultérieurement.") nous empêchait de la sauvegarder... Des tests sur PC et Macintosh avec des navigateurs différents ont, hélas, donné les mêmes résultats...

Bref, hormis ce souci passager (du moins peut-on l'espérer...), l'outil est bien fait est rapide bien que largement moins "customisable" (couleurs, pages de résultats, etc.) que ses concurrents. Mais il peut très bien convenir à une utilisation personnelle de l'outil (à noter que les macros sont également partageables entre plusieurs utilisateurs ce qui peut avoir un intérêt certain sur un intranet par exemple...).

Les Swickis d'Eurekster (<http://www.eurekster.com/>)

Nous avons déjà parlé dans cette lettre (juillet-août 2006) des Swickis d'Eurekster aussi nous n'y reviendrons pas plus... Le concept est identique à celui de ses principaux concurrents (rappelons qu'il a été l'un des premiers à mettre en place ce type d'outil) avec une interface permettant de créer de façon très complète son propre moteur de recherche vertical :



A noter que, chez Eurekster, les possibilités de configuration sont excellentes : liste d'urls à inclure ou à exclure, thématiques générales, mots clés, filtre sur les contenus adultes, recherches multimedia, ajout de nuages de requêtes, etc. La technologie de recherche utilisée est celle de Yahoo! (*Yahoo! Search Technology*).

Le système de Swikis d'Eurekster sera certainement à privilégier à Yahoo! Search Builder si l'on désire un outil personnalisé plus "pointu" que celui proposé par le portail américain. Il est surtout, en effet, remarquable de par le nombre d'options de configuration qu'il propose, au détriment d'une interface peut-être plus complexe à déchiffrer...

Rollyo (<http://www.rollyo.com/>)

Enfin, dernier outil de notre vaste tour d'horizon de moteurs personnalisables, Rollyo, l'un des plus anciens également du domaine, est bâti sur le même concept de création de votre propre util de recherche grâce à une interface simple et claire :

Peu de choix ici mais le principal est là pour créer ses propres "Searchrolls" : lister des sites web sur lesquels porteront les réponses, choisir une catégorie, proposer des mots clés, etc. Là encore, la technologie de recherche prise en compte est celle de Yahoo!, tout comme pour Eurekster. Le "Searchroll" ainsi créé viendra ensuite s'ajouter à la liste de ceux créés par défaut sur la page d'accueil de l'outil. Une option (disponible pour les utilisateurs référencés) permet de rendre public ou privé un moteur personnalisé. Un service simple et "basique" mais qui peut amplement suffire selon les besoins...

Conclusion

En quelques mois, les outils de création de moteurs de recherche personnalisés ont fleuri sur le Web, de la part de *start-ups* pionnières (Eurekster, Rollyo), rapidement suivies par les trois grands de la recherche d'information. Au-delà de l'aspect "création de moteurs thématiques pour le Web" dans le cadre de la mise en place d'un site public, un des grands intérêts de ces outils, pour nous, est bien la possibilité, en quelques minutes, de créer, au gré des besoins et des projets, des outils utilisables "en solo" ou sein d'un service, voire d'une entreprise, pour effectuer rapidement des recherches au sein d'une thématique donnée. Des moteurs "verticaux éphémères" en quelque sorte... Longtemps attendus, ce type d'outil permet aujourd'hui d'aller bien plus vite dans ses recherches et sa veille, même si elle est épisodique. Ce n'est pas là le moindre de leurs avantages... A vous de tester ces cinq outils et d'utiliser celui ou ceux qui répondent le mieux à vos attentes !