

Les outils de clusterisation des résultats de recherche (2ème partie)

[Retour au sommaire de la lettre](#)

On peut reprocher aux moteurs de recherche actuels de nous proposer des résultats uniquement sous la forme d'une simple liste de dix liens consécutifs, de façon linéaire et généraliste. Pourtant, des solutions existent pour insérer plus d'"intelligence" dans la façon dont ces résultats sont affichés. C'est le cas des systèmes de clusterisation qui permettent de classer les liens dans des dossiers thématiques qu'il est possible d'explorer pour n'obtenir qu'une partie thématisée des résultats renvoyés. Dans cet article, nous explorons de nouveaux outils qui ont choisi cette voie pour afficher leurs liens. Certains sont assez étonnants et efficaces, et tous ajoutent une nouvelle dimension à la recherche d'information sur le Web...

Il existe un reproche que l'on peut faire à beaucoup de moteurs de recherche : leurs pages de résultats sont présentées à la "queue-leu-leu" sans que vous puissiez en appréhender les différentes catégories auxquelles elles pourraient appartenir. Et pourtant, il est possible d'imaginer toutes sortes de classifications possibles comme les différentes thématiques, les types de domaines trouvés, les occurrences de tel ou tel mot-clé, etc.

Après une première partie, le mois dernier, qui explorait de nombreux outils souvent méconnus ayant choisi cette voie d'affichage des résultats renvoyés à l'internaute, nous continuons ce mois-ci notre exploration avec les métamoteurs en ligne...

Les métamoteurs (dissidents)

Dans cette partie, nous nous intéresserons aux métamoteurs qui s'appuient sur d'autres moteurs que Google.

Quintura

Nous aimons bien ce moteur qui exploite une notion appelée "nuages de tags" ou "nuage heuristique" et qui est basé sur Yahoo!. L'adresse est celle-ci : <http://www.quintura.com>. Il faut parfois se montrer patient car ce service est globalement assez lent.

Saisissez une recherche comme celle-ci : [clustering](#).

Dans le volet de gauche, vous obtenez votre premier "Tag cloud".

The screenshot shows a search results page for the query "clustering". On the left, there is a "Tag cloud" with various related terms. The most prominent term is "wclustering" in red. Other terms include "algorithms", "server", "software", "analysis", "services", "support", "windows", "linux", "hpc", "architecture", "solution", "microsoft", "storage", "topic", "high availability", "results", and "hierarchical". On the right, there is a list of search results. The top result is "Clustering - Wikipedia, the free encyclopedia" with a size of 15Kb. The second result is "Cluster analysis - Wikipedia, the free encyclopedia" with a size of 67Kb. The third result is "What is clustering? - A Word Definition From the Webopedia Computer Dictionary" with a size of 49Kb. The fourth result is "clustering" with a size of 3Kb. The fifth result is "Vivisimo, Inc - Enterprise Search, Federated Search and Clustering" with a size of 28Kb. At the top right of the results area, it says "Results found — 63600000".

Laissez le curseur de la souris sur un des mots-clés visibles pour afficher les thèmes similaires. La page de droite va automatiquement être réorganisée en fonction de la grappe qui apparaît. Vous pouvez supprimer une des occurrences visibles en cliquant sur la petite icône orange représentant une croix.

En haut à droite de la fenêtre, le lien *Settings* vous permet de personnaliser l'apparence du nuage. Les boutons *Embed* et *Share* vous permettent de partager le nuage de liens ou la page des résultats.

Clusty

Clusty (<http://clusty.com/>) est la vitrine de la société Vivisimo, spécialisée dans les technologies de clustering. Il existe des outils pour les webmasters à partir de cette page : <http://clusty.com/labs>. Clusty est également accessible à partir des mobiles en utilisant cette adresse : <http://m.clusty.com/>.

Les moteurs de recherches utilisés sont accessibles en cliquant sur le lien *détails* en haut et à droite de la page de résultats.

La barre de menu offre ce qu'on pourrait appeler des fonctionnalités de recherche horizontale... À partir d'un même mot-clé, vous pouvez le décliner en différents départements : Web, news, Wikipédia, blogs, etc.

Seul le module *Images* ne propose de clusterisation des résultats à moins de cliquer sur l'onglet *name*. Une autre remarque peut être faite : les résultats renvoyés diffèrent nettement de ceux des autres moteurs.

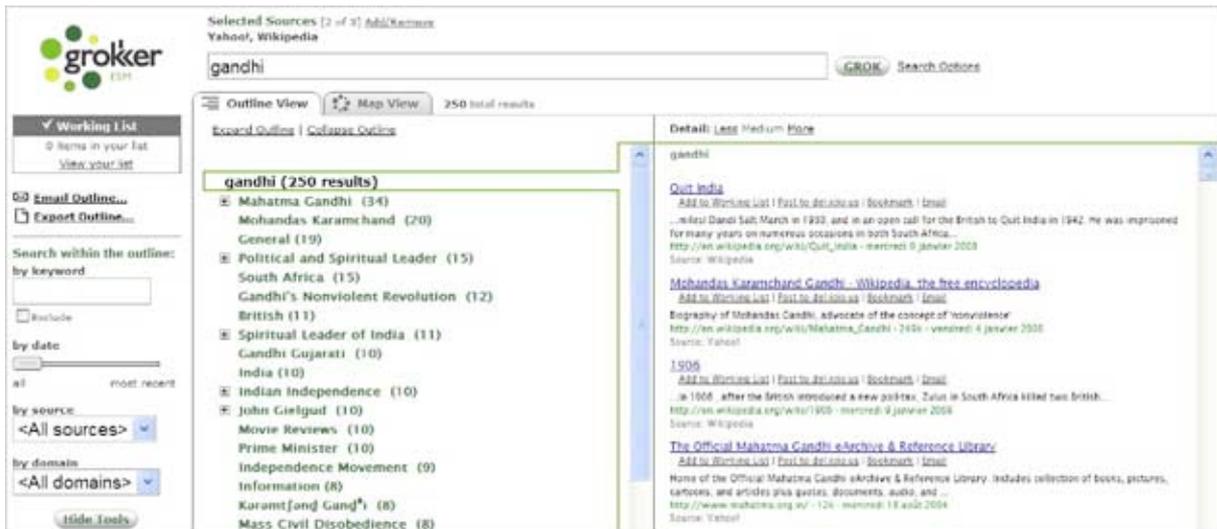
Dans tous les cas vous pouvez opérer un tri par sources et par sites.

Nous avons trouvé la clusterisation des résultats par Clusty très claire et particulièrement bien faite. Vous pouvez aussi lancer une recherche dans les clusters qui ont été détectés même s'ils ne sont pas tous visibles.

The screenshot shows the Clusty search engine interface. At the top, there is a navigation bar with links for 'web', 'news', 'images', 'wikipedia', 'blogs', 'jobs', and 'more'. The search bar contains the text 'gandhi' and a 'Search' button. Below the search bar, there are links for 'advanced preferences'. The main content area displays 'Top 234 results of at least 6,472,442 retrieved for the query **gandhi** (definition) (details)'. On the left side, there is a sidebar with 'clusters' and 'sources sites' tabs. Under 'clusters', there is a list of categories with counts: 'All Results (234)', 'Mahatma Gandhi (82)', 'Mohandas (36)', 'Indira Gandhi (23)', 'Photographs (19)', 'Gandhi Institute (16)', 'Movie (19)', 'Rajiv Gandhi (15)', 'Sonia Gandhi (14)', 'Quotes (10)', and 'India's Gandhi (8)'. Below this list is a search box for 'find in clusters' with the text 'independence' and a 'Find' button. The main results area shows a list of four items:

- WELCOME TO MAHATMA GANDHI ONE SPOT COMPLETE INFORMATION WEBSITE** - Includes chronologies, articles and links to biographies and an autobiography. www.mkgandhi.org - [cache] - Live, Gigablast, Ask
- Gandhi (1982)** - Plot Outline: Biography of Mahatma Gandhi, the lawyer who became the famed leader of the Indian revolts against the British through his philosophy of non-violent protest. www.imdb.com/title/tt0083987 - [cache] - Live, Gigablast, Ask
- Home** - The Mahatma Gandhi Canadian Foundation. For World Peace. Welcome to the Mahatma Gandhi Canadian Foundation for World ... Canada) proclaimed the week of Gandhi's birth anniversary Mahatma Gandhi Peace Week. This has since become an annual ... www.gandhi.ca - [cache] - Gigablast, Open Directory, Ask
- Manas: History and Politics, Mahatma Gandhi** - M ohandas Karamchand Gandhi was born in the town of Porbander in the state of what is now Gujarat on 2 October 1869. He had his schooling in nearby Rajkot, where ... www.sscnet.ucla.edu/southasia/History/Gandhi/gandhi.html - [cache] - Live, Gigablast, Ask

Grokker : une autre étoile montante dans le monde des recherches clusterisées ? Ce métamoteur s'appuie sur les résultats renvoyés par Yahoo!, Wikipédia et Amazon Books. Voici son adresse : <http://www.grokker.com>. C'est assez lent... Saisissez une recherche puis cliquez sur le bouton GROK...



Le lien *Search Options* vous permet de définir le nombre de résultats présentés et de les classer en fonction de leur source. Le service est en anglais mais la catégorisation des résultats est extrêmement pertinente.

Le volet de gauche vous permet d'effectuer des recherches, de filtrer les liens affichés par date, par source ou par type de domaines...

Celui de droite vous donne la possibilité de définir le niveau de détails des liens.

Cliquez ensuite sur l'onglet *Map View*.

Plus un cercle qui entoure une catégorie est important, plus il contient de nombre de résultats.

Il existe deux niveaux de zoom qui s'activent en cliquant simplement sur une des catégories.

Cliquez ensuite sur les boutons *Zoom back* et *Top* afin de revenir au point de vue initiale. Dans le volet de droite s'affichera la page concernée. On peut vraiment penser à un effet de loupe grossissante. Selon une recette apparemment bien rodée, il est aussi possible d'exporter une carte

ou de l'envoyer par mail à un de vos correspondants.

Kvisu est le nouveau moteur de recherche de Kartoo qui reprend les fonctionnalités de leur solution logicielle proposée aux entreprises (*voir l'interview de Laurent Baleydiér, DG de Kartoo, dans cette lettre*). Il propose une forme de clusterisation visuelle des résultats. Voici son adresse : <http://beta.kvisu.com>.

Vous pouvez filtrer les résultats par type, langue et date, mais aussi les classer par votes, date et pertinence.

Le volet de droite offre une catégorisation des résultats qui ressemble à des plaques tectoniques ou, plus justement, à une feuille d'un arbre parcourue de nervures. Cliquez sur une des plaques et les résultats affichés dans le cadre central vont automatiquement être réorganisés. Effet bœuf, garanti !

De manière similaire, vous pouvez développer la rubrique *Mots-clés de recherche* afin d'obtenir des classements différents.

Prenons un exemple avec cette requête : [gandhi](#).

The screenshot shows the KartooVisu search engine interface. At the top, there is a search bar with the text 'gandhi' and a 'Rechercher' button. Below the search bar, there are navigation tabs for 'Recherche', 'Résultats', 'Historique', and 'Aide'. The main content area is divided into two sections: 'Recherche gandhi' on the left and 'Cartographie thématique' on the right. The search results section displays a list of results, including 'Gandhi sur CanalPlay', 'Le Figaro Madame : Biographie', and 'Gandhi - Wikipédia'. The thematic map section shows a network of green nodes connected by lines, representing the relationships between different concepts related to Gandhi. The nodes include terms like 'non-violence', 'biography', 'histoire', 'mahatma', 'europe', 'india', 'mahatma gandhi', 'mohandas', and 'mohandas gandhi'. Below the map, there is a 'Mots-clés de recherche' section.

Nous obtenons ces plaques : *non-violence, biography, histoire, mahatma, europe, etc.*
D'une certaine manière, nous avons affiché de manière rapide les principaux concepts entourant le nom du fondateur de l'Inde moderne avec, à chaque fois, les pages web qui insistent plus particulièrement sur tel ou tel point de la vie de Gandhi, sa pensée philosophique, les questions de géopolitique, etc.

L'onglet *Historique* permet d'afficher la liste de vos recherches. Une manière pratique de se retourner sur le chemin parcouru ! C'est, à notre sens, ce que nous vu de plus abouti dans le domaine !
Pour info, Laurent Baleydiér de la société Kartoo nous a précisé qu'une version développée en Ajax serait disponible très prochainement. Restez-donc à l'écoute...

On retrouve le même concept sur l'autre moteur de cette même société mais présenté sous forme de cartes. **Kartoo** s'appuie sur les résultats délivrés par un grand nombre de moteurs dont la liste est visible à partir de cette page : <http://www.kartoo.fr/fr/kartooav.html>. Son adresse est celle-ci : <http://www.kartoo.fr/>.

Saisissez tout d'abord votre recherche. Les sites trouvés sont représentés par des pages plus ou moins grosses, selon leur pertinence. Ce moteur utilise un certain nombre de conventions visuelles dont la liste complète est visible à cette adresse : <http://www.kartoo.net/a/fr/aide04.html>.
Les liens qui unissent les différentes pages vont être affichés. Sélectionnez un site. Dans le volet de gauche, il est possible :

- D'ouvrir la page correspondante ;
- D'afficher les sites semblables ;
- D'afficher la carte de ce site ;
- D'atteindre la page d'accueil du site ;
- D'effacer la page trouvée de la carte ;
- De copier ce site.

Cette dernière option est utile si vous voulez créer une nouvelle carte à partir d'éléments récupérés sur une précédente carte.

Notez qu'il est possible de réorganiser les pages qui sont visibles sur une carte en les déplaçant à l'aide du curseur de la souris. On parle alors de graphes sémantiques... Le schéma (quelque peu brumeux) qui est visible en arrière-plan sera modifié en temps réel.

Vous pouvez aussi préciser votre recherche de départ en sélectionnant un ensemble de mots-clés qui ont été détectés. Si vous gardez le curseur de la souris sur un de ces mots-clés, le lien qui relie, à travers ce mot-clé, les sites trouvés va s'afficher.

En cliquant dessus, notre recherche peut alors devenir celle-ci : [gandhi "january 1948"](#). Il est ensuite possible de filtrer les résultats en sélectionnant un mot-clé puis en cliquant sur une de ces commandes : *Ajouter à la recherche*, *Enlever de la recherche*, *Effacer de la carte*, etc. Mais la présence d'un encart présentant des liens sponsorisés rend ces possibilités extrêmement compliquées à utiliser...

À gauche de la barre de recherche, les différents boutons vous permettent :

- D'afficher l'historique de vos recherches ;
- De définir les options de recherche ;
- De créer une nouvelle carte ;
- D'enregistrer une carte ;
- D'imprimer la carte.

Il existe aussi une version HTML de ce métamoteur : <http://www.kartoo.fr/fr/servlet/H>.

Un autre moteur de clusterisation appelé **Ujiko** (et conçu aussi par Kartoo) est accessible à partir de cette adresse : <http://www.ujiko.com/v2a/flash.php?langue=fr>... L'interface tient plus de l'OVNI que d'autre chose mais le moteur est très agréable à utiliser.



Article écrit par un rédacteur free-lance pour la lettre "Recherche & Référencement"

Réagissez à cet article sur le blog des abonnés d'Abondance :
<http://abonnes.abondance.com/blogpro/2008/02/les-outils-de-clusterisation-des.html>