

Chunkit et SurfCanyon : deux outils pour améliorer les résultats des grands moteurs de recherche

[Retour au sommaire de la lettre](#)

Les deux outils présentés dans cet article, ChunkIt et SurfCanyon, s'appuient sur les moteurs de recherche traditionnels (Google, Yahoo!, Live...) pour tenter de leur apporter des fonctionnalités supplémentaires afin de nous aider à mieux chercher l'information sur le Web au quotidien. L'un explore le contenu des pages web proposées comme résultat par les moteurs, l'autre tente d'identifier d'autres pages pertinentes pour un même site web. Deux utilitaires à tester et qui pourraient vite devenir indispensables dans votre trousse à outil de chercheur web...

D'opérateurs booléens en fonctionnalités de filtrage avancées, on croyait avoir épuisé les ressources de nos moteurs de recherche favoris. Pourtant, à coup de traitement automatique du langage et d'amélioration de la lisibilité, deux sociétés viennent de lancer des outils permettant d'exploiter encore plus efficacement leurs résultats.

ChunkIt, pour aller à l'essentiel

ChunkIt, de la société TigerLogic (<http://www.tigerlogic.com/ChunkIt/home.html>) se présente comme une barre d'outils à installer dans votre navigateur (Firefox ou IE7). Elle propose deux fonctionnalités intéressantes :

- **Améliorer les résultats de recherche des moteurs classiques.** ChunkIt fonctionne à partir des résultats de plusieurs moteurs pris un par un (ce n'est donc pas un métamoteur). Il peut s'agir de Google, de Yahoo!, de LiveSearch, d'AOL ou bien encore d'Ask, ce choix s'effectuant au moment de l'interrogation. Vous obtenez alors une page de résultats de ce type :

The screenshot displays the ChunkIt search results page. At the top, there's a search bar containing the text "Intelligence économique". Below the search bar, the results are organized into a list. The first result is "Blogs et Intelligence économique" from "http://blogsette.blogspot.com". It includes a brief description: "Ca parle de quoi? challenges, espionnage, Intelligence Economique, métier en or, salaire, Veille". The second result is "Egideria" from "http://www.egideria.com", with the description "Welcome - Our Mission ...". The third result is "Le blog de l'intelligence économique - Blogs - Les Echos" from "http://blogs.lesechos.fr", with the description "Mettre l'intelligence économique au service des entreprises, 'sensibles' ou pas. Telle l'ambition de ce blog réalisé avec les experts de 'l'Académie de ...". Each result has a "Chunk Links" button and an "Email Chunks" button. On the left side, there's a sidebar with a list of results and a "No relevant content found at this location." message. At the bottom, there's a search bar with the text "Intelligence économique" and a "Search" button. The interface also includes a footer with links for "Search within results", "Language Tools", "Search Tips", "Dissatisfied? Help us improve", and "Try Google Experimental".

Ici, chaque page de résultat proposée par Google a été visitée par ChunkIt qui en a tiré les éléments les plus significatifs et replace vos mots-clés dans leur contexte (à gauche), un bon moyen pour ne plus perdre de temps à visiter des pages non pertinentes.

Si vous cliquez sur l'une de ces icônes , vous faites apparaître l'extrait surligné dans son contexte sur la page d'origine.

9. [Le blog de l'intelligence économique - Blogs - Les Echos](#)
<http://blogs.lesechos.fr>
[Chunk Links](#) [Hide Chunks](#) [Email Chunks](#)

Les récentes affaires qui ont présenté l'intelligence économique sous un jour négatif, rendent encore plus difficile l'implantation d'une dynamique d'IE dans une entreprise. Comment un chef d'entreprise convaincu du bienfait de l'IE peut-il la promouvoir sans éveiller les interrogations d'au moins une partie de ses salariés ? Premier élément de réponse, il n'est pas nécessaire d'utiliser l'expression « **intelligence économique** » pour faire de « l'intelligence économique ». Cette discipline (...)

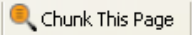
aspect que le mimétisme économique - une plaie (...)
[Lire la suite](#) - [1 commentaire\(s\)](#)

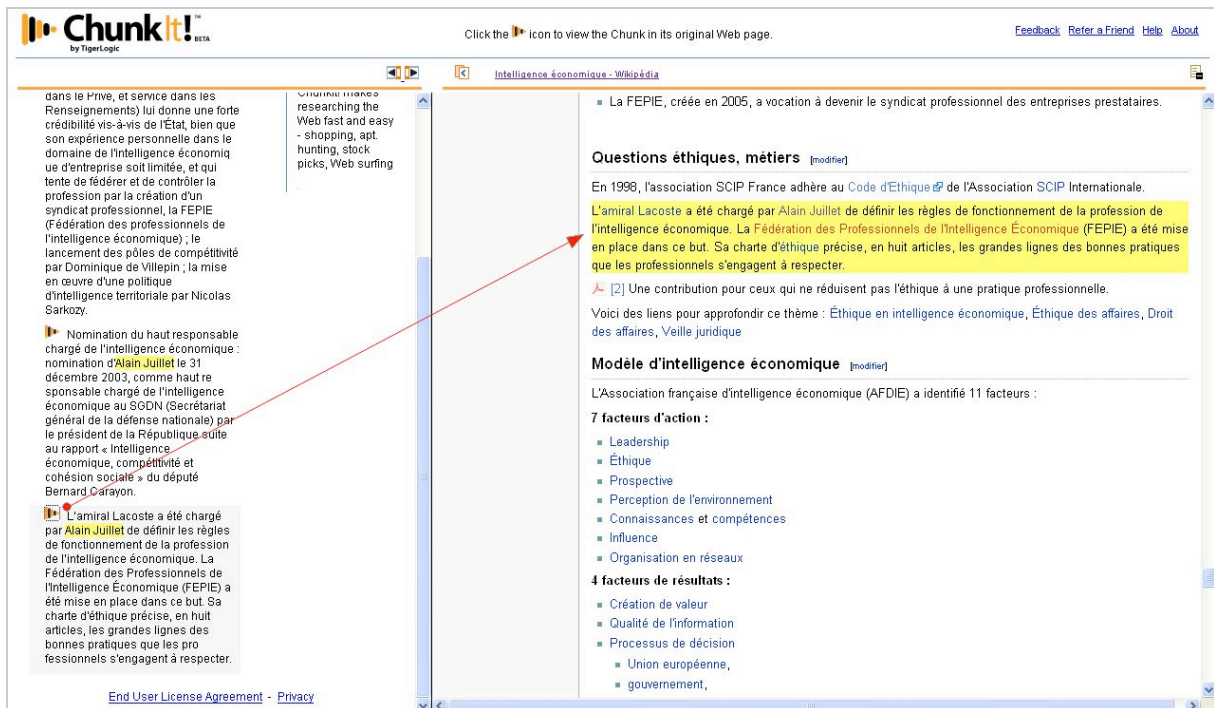
Comment implanter une culture de l'Intelligence Economique ?
Jérôme Bondu - [06/05/2008]

Les récentes affaires qui ont présenté l'intelligence économique sous un jour négatif, rendent encore plus difficile l'implantation d'une dynamique d'IE dans une entreprise. Comment un chef d'entreprise convaincu du bienfait de l'IE peut-il la promouvoir sans éveiller les interrogations d'au moins une partie de ses salariés ? Premier élément de réponse, il n'est pas nécessaire d'utiliser l'expression « **intelligence économique** » pour faire de « l'intelligence économique ». Cette discipline (...)
[Lire la suite](#) - [6 commentaire\(s\)](#)

Vous avez ainsi la possibilité d'aller directement à ce qui vous intéresse.


- **Explorer le contenu d'une page web.** Nous utilisons tous la commande Ctrl+F qui permet de rechercher un mot-clé dans la page que vous avez à l'écran. ChunkIt permet désormais d'aller beaucoup plus loin avec cette fonctionnalité. Lorsque vous visitez une page avec beaucoup de contenu et souhaitez trouver un mot-clé particulier il vous suffit de le taper dans la barre d'outils

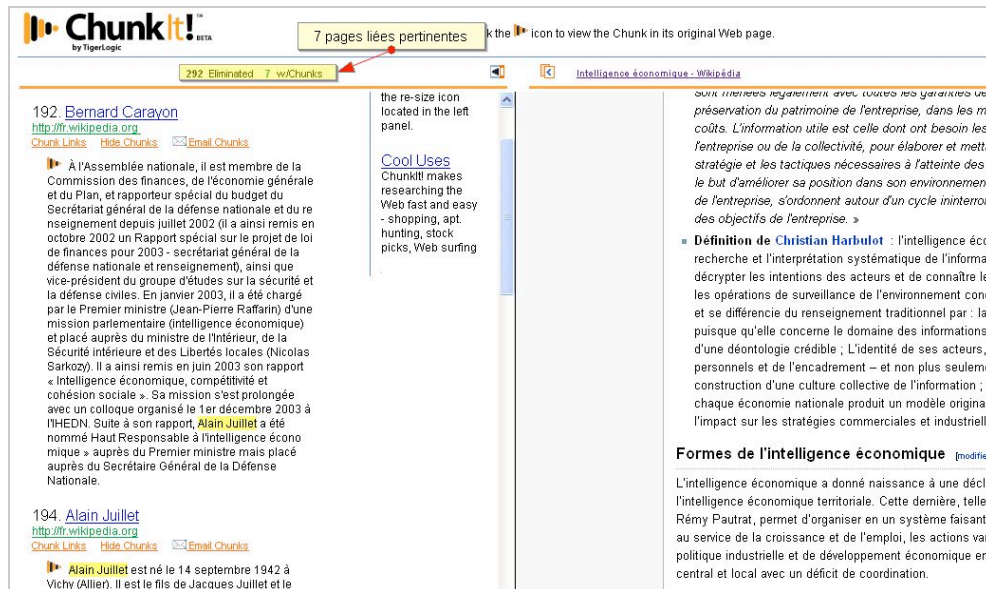
puis de cliquer sur  **Chunk This Page**. Si par exemple vous vous trouvez sur la page Wikipedia consacrée à l'"Intelligence économique" et souhaitez lire ce qui concerne Alain Juillet, il vous suffit de taper son nom dans ChunkIt pour obtenir la page de résultats suivante :



The screenshot shows the ChunkIt interface. On the left, a sidebar lists search results for 'Alain Juillet'. The main content area displays the Wikipedia article 'Intelligence économique' with a red arrow pointing to a highlighted section about 'Alain Juillet'. The right sidebar contains two sections: 'Questions éthiques, métiers' and 'Modèle d'intelligence économique', both with expandable content.

Chaque occurrence du mot recherché apparaît à gauche surlignée et remise dans son contexte. En cliquant sur l'icône ChunkIt vous faites apparaître dans la page de droite le même extrait surligné dans la page originale. Il vous suffit alors de naviguer d'extrait en extrait pour ne lire que ceux qui vous intéressent.

Mais ChunkIt va plus loin. Si maintenant vous cliquez sur  **Chunk Right-Page Links** l'outil va alors suivre chacun des liens présents dans votre page initiale et vous présenter des extraits de toutes les pages correspondantes dans lesquelles votre mot-clé apparaît. C'est comme si vous affiniez votre requête en interrogeant une sélection de pages potentiellement pertinentes puisque rattachées à votre page de démarrage par un lien hypertexte. Dans notre cas nous obtenons sept autres pages liées citant le nom d'Alain Juillet dont l'article Wikipedia le concernant :



A noter que ChunkIt permet aussi :

- de faire une recherche sur les mots de même famille (lemmatisation)
- d'envoyer par email une page de résultats structurée

Il est pour l'instant gratuit dans sa version bêta. Il faut s'inscrire sur le site et demander une invitation pour pouvoir le tester. A votre tour...

SurfCanyon fouille les pages de résultat en profondeur

Encore moins intrusif, Surf Canyon (<http://www.surfcanyon.com/>) est une extension à ajouter à votre navigateur (Firefox, IE7) qui améliore la recherche d'informations sur les principaux moteurs généralistes (Google, Yahoo!, LiveSearch) de manière significative.

Depuis Google, tous les moteurs proposent des listes de résultats basées sur la popularité des pages web qu'ils indexent. Ils se distinguent les uns des autres par la manière dont ils calculent cette popularité mais se ressemblent tous en un point : leurs résultats peuvent s'étendre sur plusieurs dizaines de pages et rares sont les internautes qui dépasseront la 3ème ou la 4ème. Résultat : nous utilisons les pages les plus connues de tous mais manquons celles qui sont pertinentes et pas encore suffisamment populaires pour être en tête de classement. Découragés, nous passons parfois sans le savoir à un clic de la page qui répondait parfaitement à notre question et nous en sommes conscients ! Frustrant.

En mêlant le traitement sémantique de l'ensemble des résultats d'une requête et l'analyse de vos clics en temps réel, SurfCanyon vous permet de découvrir et d'exploiter ces pages profondes. Pour l'utiliser il suffit d'interroger votre moteur de prédilection, au hasard Google, comme vous le faites habituellement. Vous obtenez alors une page classique dans laquelle chaque résultat est suivi d'une icône en forme de cible :



En cliquant sur l'une de ces icônes, vous donnez à SurfCanyon des indices sur ce qui vous intéresse vraiment. Si dans notre exemple vous cliquez sur le second résultat le système va "matcher" le contenu de la page correspondante avec celui des autres pages de résultats et vous recommander les trois dont le contenu est le plus proche :

Définition de la **veille stratégique** 

La **veille stratégique** est le processus informationnel volontariste par lequel l'entreprise recherche des informations à caractère anticipatif concernant ...
www.sup.adc.education.fr/bib/publ/guides/veille/definition.htm - 11k -
[En cache](#) - [Pages similaires](#) - [À noter](#)

☒ Surf Canyon recommends 3 search results:

Veille stratégique (from page 2) 

Publications : Guides. Veille stratégique. Concepts et démarche de mise en place dans l'entreprise. Par : Humbert Lesca · L'Auteur · Sommaire · Commentaire ...
www.sup.adc.education.fr/bib/publ/guides/veille/ - 4k - [En cache](#) - [Pages similaires](#) - [À noter](#)

Etude : La **Veille Stratégique** - Les Yeux et les Oreilles de Votre ... (from page 2) 

Processus par lequel l'entreprise observe, recherche, localise, traite, analyse, et diffuse l'Information Stratégique à des buts Décisionnels et ...
www.3ie.fr/.../etude/etude__La_Veille_Strategique_Les_Yeux_et_les_Oreilles_de_Votre_Entreprise.htm - 12k -
[En cache](#) - [Pages similaires](#) - [À noter](#)


La **Veille Stratégique** (from page 9) 

La **Veille Stratégique** est le processus par lequel l'entreprise observe, recherche, localise, traite, analyse, et diffuse l'Information Stratégique à des ...
erwan.neau.free.fr/information_veille_strategique.htm - 19k -
[En cache](#) - [Pages similaires](#) - [À noter](#)

Classement initial de la page découverte par SurfCanyon


On voit dans cet exemple que si les deux premiers résultats étaient présents dès la seconde page de résultats, le troisième se trouvait sur la 9ème...

Si vous n'êtes pas satisfait des réponses vous pouvez encore affiner ces résultats selon la même principe avec un second puis un troisième niveau de filtrage :


Définition de la **veille stratégique** 

La **veille stratégique** est le processus informationnel volontariste par lequel l'entreprise recherche des informations à caractère anticipatif concernant ...
www.sup.adc.education.fr/bib/publ/guides/veille/definition.htm - 11k -
[En cache](#) - [Pages similaires](#) - [À noter](#)

☒ Surf Canyon recommends 3 search results:


Veille stratégique (from page 2) 

Publications : Guides. Veille stratégique. Concepts et démarche de mise en place dans l'entreprise. Par : Humbert Lesca · L'Auteur · Sommaire · Commentaire ...
www.sup.adc.education.fr/bib/publ/guides/veille/ - 4k - [En cache](#) - [Pages similaires](#) - [À noter](#)

Etude : La **Veille Stratégique** - Les Yeux et les Oreilles de Votre ... (from page 2) 

Processus par lequel l'entreprise observe, recherche, localise, traite, analyse, et diffuse l'Information Stratégique à des buts Décisionnels et ...
www.3ie.fr/.../etude/etude__La_Veille_Strategique_Les_Yeux_et_les_Oreilles_de_Votre_Entreprise.htm - 12k -
[En cache](#) - [Pages similaires](#) - [À noter](#)

☒ Surf Canyon recommends 3 search results:

01net. - De la **veille stratégique** à la sécurité de l'information (from page 2) 

L'intelligence économique n'est pas l'apanage des multinationales. Elle concerne en réalité toutes les entreprises, PME comprises.
www.01net.com/article/273356.html - 112k - [En cache](#) - [Pages similaires](#) - [À noter](#)

☒ Surf Canyon recommends 3 search results:

Accueil - Intelink : Actualités

Intelink, intelligence et système de **veille stratégique**.
www.intelink.info/ - 51k - [En cache](#) - [Pages similaires](#) - [À noter](#)

De fait plus vous cliquerez sur les cibles et plus Surf Canyon disposera d'éléments lui permettant de comprendre finement ce que vous recherchez.

A l'usage, Surf Canyon s'avère vite indispensable même si l'on peut regretter qu'il ne dispose pas (encore) d'un bouton de "débrayage" pour revenir à une page de résultats classique lorsque cela s'avère nécessaire.

Détail d'importance, Surf Canyon, a été optimisé pour la langue française (son créateur a passé plusieurs années chez Hewlett-Packard à Grenoble).

Quoiqu'il en soit, ChunkIt et Surf Canyon ouvrent la voie à une nouvelle génération d'outils qui permettra de toujours mieux exploiter les ressources infinies indexées par les grands moteurs de recherche. Ceux-ci ont beau disposer de fonctionnalités avancées, la relative complexité de leur mise en oeuvre n'incite pas les internautes à les utiliser et ils restent alors souvent sur des expériences frustrantes. Dans ce contexte, des outils peu intrusifs et "user-friendly" tels que ceux présentés ici ont une réelle carte à jouer car de fait, même si les utilisateurs finaux ne comprennent pas bien comment Surf Canyon traite les pages (secret de fabrique oblige), ils obtiennent des résultats plus pertinents qu'auparavant. N'est-ce pas l'essentiel à leurs yeux ?

Christophe Deschamps

Consultant et formateur en gestion de l'information.

Responsable du blog Outils Froids (<http://www.outilsfroids.net/>)

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

<http://abonnes.abondance.com/blogpro/2008/06/chunkit-et-surfcanyon-deux-outils-pour.html>