

Comment intégrer un moteur à votre site : Fast

Nous vous proposons, dans cette série d'articles, les différentes façons d'insérer dans vos pages les moteurs leaders de la recherche d'information sur le Web : Google, Fast, Inktomi, Voila, AltaVista, etc. Chacun de ces outils propose, de façon plus ou moins pointue, plus ou moins gratuite aussi, d'insérer un moteur de recherche web dans vos pages. Le but de ces articles est de vous faire (re- ?) découvrir les différentes manières d'effectuer cette tâche et les différentes offres proposées par les moteurs pour ce faire.



Après Google le mois dernier, voici les offres de Fast, son actuel challenger, passées au crible.

2 offres possibles sur Fast

L'offre de Fast en la matière est ultra-simple : une possibilité gratuite et une payante. En d'autres termes, pourquoi faire simple quand on peut faire compliqué ;-) ?

L'offre gratuite est réduite à sa plus simple expression : il s'agit de l'ajout d'un simple formulaire de recherche (sur AllTheWeb, le site "labo de recherche" de Fast) dans vos pages. L'offre est présentée à l'adresse :

http://www.alltheweb.com/help/webmaster/add_to_site.html

La procédure est la suivante : vous choisissez votre type de recherche (Web, multimedia, FTP, actualité), l'encodage de vos pages (selon votre pays), et l'outil vous donne le code HTML à insérer dans vos documents Web.

Si le formulaire est légèrement configurable en fonction de vos besoins, en revanche la page de résultats sera celle d'AllTheWeb. Aucune possibilité de configuration des pages-résultats à votre charte graphique n'est disponible ici. Bref, une possibilité similaire au "Link to Google" (<http://www.google.com/searchcode.html>) vu le mois dernier.

Côté "offre gratuite" donc, peu de nouveau et rien de très original. Pour obtenir de réelles et complètes possibilités de configuration, il faudra donc se tourner vers l'offre "Fast Web Search" (<http://www.fastsearch.com/products/websearch/index.asp>), payante mais qui propose de nombreuses fonctionnalités pointues et professionnelles pour mettre Fast "à vos couleurs".

Si elle n'est pas à la portée de toutes les bourses (voir plus loin), l'offre est, ici très intéressante au niveau de la configuration des pages-résultats. De nombreuses informations m'ont été fournies par Fast à ce sujet (merci à eux), ce qui va nous permettre de mieux comprendre, sur cet exemple, comment interfacier un site web avec un moteur de recherche leader. Finalement, il n'y a rien de très compliqué techniquement, si ce n'est que ces possibilités sont réservées aux développeurs...

La requête

Tout d'abord la première étape, après que vous ayez installé un formulaire de recherche sur vos pages : le visiteur tape un mot clé et clique sur le bouton de soumission. Vous devez donc, en tant que site client de Fast, envoyer cette requête au moteur. Cela se fait sous la forme d'une url comprenant un certain nombre de paramètres. En voici la liste :

Champ	Format	Commentaires
query	texte	La requête, le(s) mot(s) clé(s) demandé(s).
type	texte. Valeur : all (par défaut), any, phrase	Et, OU ou l'expression (guillemets).

hits	entier (par défaut 10)	Nombre maximal de résultats par page.
offset	entier (par défaut 0)	Rang du premier résultat (première page : offset=0, deuxième page : offset= 10 - avec paramètre "hits"= 10 -), etc.
language	texte	Langue utilisée, sur deux caractères (fr=français).
encoding	texte	Encodage utilisé (UTF-8 ou autre).
collapse	booléen	Clustering ou non (1 seul lien par site web ou tous les liens d'un site répondant à la requête).
qtf_suggestonly	booléen	Présence ou non de suggestions permettant d'affiner sa requête. Voir par exemple, sur AllTheWeb, la page de résultats pour une requête comme "linux device driver" (sans guillemets) et la zone "About your query" à droite.
qtf_protectquery	booléen	Présence ou non de suggestions de requêtes permettant d'affiner sa requête, sous la forme de liens "prêts à cliquer". Exemple : voir la section "Narrow your search" sur la droite de certaines pages-résultats d'AllTheWeb.

Voici donc, à la lumière des paramètres précédents, un exemple de requête envoyée directement à Fast suite à une saisie sur votre site de la part d'un visiteur :

http://***/std_xml_d00?query=linux+device+drivers&type=all&language=fr&hits=20&offset=100

En règle générale, la mention ******* est remplacée par un nom de serveur ou un numéro IP fourni par Fast à ses clients.

La mention "std_xml_d00" précise le format sous lequel on désire recevoir la réponse. Plusieurs formats sont possibles, selon le niveau de précision que l'on désire recevoir en retour (simple "fiche de présentation" des liens à afficher ou informations complémentaires concernant la date de dernière modification, le poids, la langue, des options de clustering, etc.). Nous n'entrerons pas plus dans les détails sur ce point.

Obtentions des résultats

Vous avez "capté" la requête de votre internaute et vous l'avez envoyée à Fast. Ce dernier va donc vous renvoyer le résultat désiré sous la forme d'un fichier XML comprenant les différentes informations demandées, qui vont vous permettre de créer une page de résultats selon votre propre charte graphique et d'y insérer les données fournies par le moteur.

Nous n'étudierons pas ces résultats en détails (il faudrait un article de plus de 20 pages pour cela) mais sachez que ces données sont divisées en quatre grandes familles :

- Les informations globales sur la recherche effectuée : le rappel des infos envoyées dans la première étape - la requête - , le nombre global de résultats trouvés, le temps nécessaire pour effectuer la recherche, ainsi que d'éventuels codes d'erreur.
- Des informations de navigation : url des pages suivante/précédente, offset (voir tableau ci-dessus) du premier lien de la page, etc.
- Des informations sur les différentes possibilités de suggestions associées à la requête demandée : autres mots clés, autres urls, autres expressions, etc. Il peut s'agir également de liens vers d'autres bases de données (Encarta, Scirus, Internet Movie Database, base de données sur le football, etc.).

- Et enfin, bien sûr, les "fiches descriptives" de chacun des liens correspondant à la demande. Pour chacun de ces liens, de nombreuses informations sont disponibles, selon le degré de précision que vous désirez afficher sur vos pages : rang, titre, description, résumé, url, langue, date de dernière modification du document, taille (en octets), le score de pertinence, le nombre de liens répondant à la requête sur le même site, etc.

Pour les plus endurcis ;-) d'entre vous, voici un exemple de fichier XML renvoyé par Fast (seuls les 2 premiers liens renvoyés sont proposés ici) :

```
<?xml version = "1.0" encoding="utf-8"?>
<!DOCTYPE RESULTPAGE SYSTEM "http://***/dtds/result_d00.dtd">

<RESULTPAGE QUERYSTRING="Where can I find information about new york yankees"
QUERYTYPE="all" FIRSTRESULT="11" NUMRESULTS="10" TOTALHITS="130788"
SEARCHTIME="0.342">

<NAVIGATION>
<PREV FIRSTNUM="1"
LASTNUM="10">/std_xml_d00?query=where+can+i+find+information+about+new+york+yankee
s&amp;type=all&amp;hits=10&amp;dataset=0</PREV>
<NEXT FIRSTNUM="21"
LASTNUM="30">/std_xml_d00?query=where+can+i+find+information+about+new+york+yankee
s&amp;type=all&amp;hits=10&amp;dataset=0&amp;offset=20</NEXT>
</NAVIGATION>

<FEEDBACK>
<FEEDBACKITEM NO="0" SOURCE="FastQT_Phase" ACTION="Modified the query" MESSAGE=""
QUERY="where can I find information about &quot;new york yankees&quot;" CUSTOM="imdb_aka-
titles"/>
<FEEDBACKITEM NO="1" SOURCE="FastQT_NonPhrase" ACTION="Modified the query"
MESSAGE="" QUERY=" &quot;new york yankees&quot;" CUSTOM="where"/>
</FEEDBACK>

<RESULT NUM="11">
<TITLE>WashingtonPost.com: New York Yankees</TITLE>
<DESCRIPTION>Daily and archived news and comprehensive batting and pitching
statistics.</DESCRIPTION>
<URL>http://www.washingtonpost.com/wp-srv/sports/longterm/sportusa/ny/yankees.htm</URL>
<LANGUAGE>en</LANGUAGE>
<MODDATE>972781354</MODDATE>
<SIZE>7986</SIZE>
<SCORE>339</SCORE>
</RESULT>

<RESULT NUM="12">
<TITLE>SportingNews.com - Baseball : New York Yankees</TITLE>
<DESCRIPTION>Rosenthal | Our picks | Scoreboard | World Series History Play Ultimate Playoff
Baseball!</DESCRIPTION>
<URL>http://www.sportingnews.com/baseball/teams/yankees/</URL>
<LANGUAGE>en</LANGUAGE>
<MODDATE>1004353383</MODDATE>
<SIZE>23524</SIZE>
<SCORE>331</SCORE>
</RESULT>

</RESULTPAGE>
```

Et ainsi de suite... A vous, donc, par la suite, d'intégrer ces différents champs dans vos pages de résultats. Simple, non ? ;-)

Evidemment, si vous ne développez pas vous-même, il vous faudra travailler avec un programmeur pour traiter toutes ces informations "à votre sauce"...

Point important : tout client de fast a à sa disposition un outil appelé "Rank Tuning Toolkit". Il s'agit d'un outil très pratique qui permet de "jouer" sur les curseurs et sur les poids attribués à certains critères de pertinence pour obtenir un "ranking" différent en fonction de ses propres besoins. Le client pourra ainsi jouer sur des critères comme la langue, le filtre familial, les noms de domaine (possibilité de mettre en avant les sites ayant un nom de domaine propre par rapport aux sites perso de type multimaniamania ou autres), etc. pour proposer à ses internautes des résultats qui peuvent donc être tout à fait différents de ceux proposés par les autres clients de Fast (AllTheWeb, Lycos, etc.). Un "bonus" ("boost") peut également être attribué à certains liens, sur certaines recherches, comme dans l'exemple ci-dessous, tiré de l'utilisation de l'outil "Rank Tuning Toolkit" sur l'offre "Data Search" (recherche sur les intranets) de Fast :

The screenshot shows the 'fastdatasearch' interface. At the top, it says 'fastdatasearch™ Relevancy Administration Toolkit'. Below that, it indicates the collection is 'cars'. There are three sections for 'Modify Relevancy': 'Absolute Query Boost', 'Relative Query Boost', and 'Relative Document Boost'. Each section has a 'Specify Boost by Searching' button. Below these is a search bar with 'cars' entered and a 'Search' button. The results section shows '3 documents found.' and lists three results with their ranks and boost values. A sidebar on the right provides information about the query, including a spellcheck notice and suggested queries like 'lemmas.all:cars'.

Syntaxe d'interrogation de Fast

Les documents fournis par Fast proposent également une annexe indiquant la syntaxe d'interrogation complète de Fast. Une bonne occasion pour faire un point très complet sur ces possibilités, très nombreuses par ailleurs...

Syntaxe	Commentaires	Exemple de requête
normal.all	Recherche dans toute la page : titre, "headers (H1-H6)" et "body".	normal.all:strasbourg
normal.title	Recherche uniquement dans le titre des pages.	normal.title:strasbourg <i>Autre possibilité :</i> title:strasbourg
normal.titlehead	Recherche uniquement dans le titre et les "headers" (H1...H6) des pages.	normal.titlehead:strasbourg

normal.anchors	Recherche uniquement dans le texte des liens .	normal.anchors:référencement
url.all	Possibilité de recherche sur une url complète.	url.all: michelin
url.host	Recherche sur le nom de serveur ou le domaine.	url.host: www.michelin.com url.host: michelin.com url.host: com
url.domain	Recherche sur le nom de domaine.	url.domain: www.michelin.com url.domain: michelin.com
url.tld	Recherche uniquement sur le domaine.	url.tld: com url.tld: fr <i>Autre possibilité :</i> domain: com domain: fr
link.all	Renvoie les pages contenant un lien vers l'url indiquée.	link.all: www.abondance.com
lang.all	Langue de recherche.	lang.all: fr
properties.all	Format demandé. Peut prendre les valeurs suivantes : image, audio, mp3, video, shockwave, java, javascript, vml, acrobat, vbscript, real.	properties.all: image properties.all: mp3
size.all	Permet de de définir la taille des pages recherchées (en octets).	motclé size.all: 42000 motclé size.all: < 42000 motclé size.all: > 42000 motclé size.all: [1024; 2048]
date.all	Recherche sur la date de dernière modification du document (sous la forme du nombre de secondes depuis le 1er janvier 1970 à minuit) .	cars date.all: 972781354
fast.type	Permet d'effectuer (avec la valeur "offensive") une recherche avec filtre parental.	sex -fast.type: offensive
classification.all	Permet d'effectuer des recherche selon des critères géographiques. Exemples de valeurs possibles : geonamerica (restriction sur les sites en .com, .net, .gov et .mil), geoeurope (sites européens), geoseasia (asie du sud-est), geoindiaasia (Inde et Asie), geosamerica (Amérique du Sud), geoceania (Océanie), geoafrika (Afrique), geocanada (Canada). D'autres valeurs sont possibles. L'option permet également de réduire sa recherche uniquement aux pages d'accueil des sites avec l'option "cathomepages".	IBM classification.all: geonamerica IBM classification.all: geoindiaasia IBM classification.all: cathomepages
urldepth.all	Permet de limiter une recherche à un certain niveau d'arborescence des sites web à explorer. Exemples : urldepth.all: 0 -> http://www.foo.com/ urldepth.all: 1 -> http://www.foo.com/about.html urldepth.all: 1 -> http://www.foo.com/about/ urldepth.all: 2 ->	abondance urldepth.all: 0 abondance urldepth.all: < 2

	http://www.foo.com/about/search.html urldepth.all:2 -> http://www.foo.com/a/b/	
--	---	--

Plutôt complet comme syntaxe d'interrogation, non ? ;-)

Comparatif des solution d'interrogation du moteur Fast

Voici donc un tableau récapitulatif des deux solutions proposées par Fast pour intégrer son moteur dans vos pages :

	Coût mensuel	Coût à la requête	POUR	CONTRE
Formulaire d'interrogation	0	0	<ul style="list-style-type: none"> - Gratuit - Très simple - Très rapide d'installation - Recherche sur l'index complet de Fast 	<ul style="list-style-type: none"> - Aucune personnalisation possible
Fast Web Search	50 000 US\$	2 US\$ par tranche de 1000 requêtes	<ul style="list-style-type: none"> - Haut niveau de configuration des pages de résultats. - Propose toutes les options que l'on peut demander pour ce type d'interface. 	<ul style="list-style-type: none"> - Cher - Nécessite des développements informatiques internes au site.

Notons également, pour être le plus complet possible, qu'une solution de recherche interne sur votre site web (moteur de recherche uniquement sur les pages de votre site) est proposée par Fast, couplée à l'offre Fast PartnerSite Plus (<http://www.fastsearch.com/products/partnersite/plus.asp>). Pour avoir accès à cette offre de moteur de recherche interne, vous devez donc souscrire obligatoirement à l'offre de référencement payant à laquelle elle est couplée. Nous n'en parlerons pas plus dans le cadre de cet article. Une autre offre, intitulée "Fast Data Search" (<http://www.fastsearch.com/products/datasearch/index.asp>) existe également. Elle permet d'intégrer la technologie Fast sur un intranet d'entreprise. Là aussi, nous ne la détaillerons pas dans le cadre de cet article.

Conclusion

Fast offre donc deux niveaux de prestations pour insérer son moteur sur vos pages. L'une est gratuite mais trop "simple" pour être réellement utile. L'autre est très bien documentée, propose de nombreuses fonctionnalités de configuration, etc. Mais elle n'est à la portée que des grands portails, financièrement parlant...

Bref, on aurait bien aimé une offre intermédiaire qui mette la recherche Fast à la portée des bourses un peu moins remplies ;-). Cela aurait comme incidence immédiate une meilleure visibilité de Fast sur le Web ainsi qu'une possibilité, pour les webmasters de "petits" et "moyens" sites, de proposer une fonctionnalité "moteur de recherche" fiable et complète à peu de frais... Au fait, comment on dit "s'il vous plaît" en norvégien (car n'oublions pas que Fast est originaire de ce pays, où les équipes de R&D se trouvent encore) ??

PS : Après que nous leur ayons transmis l'idée, Fast a promis qu'il allait y réfléchir rapidement...

Ceci dit, si votre site commence déjà à générer un trafic non négligeable, il existe également une solution pour intégrer Fast dans vos pages, en passant par Espotting ou Overture. Ces deux prestataires peuvent vous permettre (si votre "dossier de candidature" en tant que site affilié est

accepté, bien sûr) d'insérer un formulaire dans vos pages avec, bien entendu, leurs liens sponsorisés. Mais vous avez également la possibilité de rajouter, après la liste de ces liens, des résultats "moteur de recherche" qui sont de plus en plus souvent ceux de Fast (mais Inktomi est également possible)... Voir un exemple avec le moteur de recherche d'Abondance à l'adresse : <http://www.abondance.com/recherche/>