

Comment intégrer un moteur à votre site : Google

Nous allons vous proposer, 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 sur votre site, dans vos pages. La plupart du temps, deux options sont possibles : une recherche "classique" sur le Web et une recherche "intra-site", uniquement sur les pages de votre site web. 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.

Chaque mois, un acteur majeur du domaine sera traité dans la lettre R&R. Pour commencer cette série d'articles, nous avons choisi de nous pencher sur le leader incontesté actuellement des moteurs de recherche : Google.

Celui-ci propose effectivement de nombreuses possibilités pour insérer son moteur sur un site web. Certaines sont très simples, d'autres plus spécifiques, plus complètes ou plus complexes. Voici la liste des différentes prestations proposées :

Installations gratuites

Google propose plusieurs possibilités pour installer son outil de façon gratuite sur un site web distant :

- Offre "**Buddy Link to Google**" (<http://www.google.com/searchcode.html>) : l'option la plus simple, proposée par tous les outils de recherche, mais aussi, et de loin, la plus limitée. Elle se concrétise sous la forme d'un code HTML à insérer dans vos pages et qui représente une "boîte de recherche" aux couleurs du moteur (voir illustration ci-dessous). Les internautes qui visitent le site proposant cette "search box" peuvent ainsi effectuer directement des recherches sur Google.



L'offre, si elle est gratuite, ne propose aucune option de personnalisation de la page de résultats (mais la boîte de recherche, elle, peut être modifiée à souhait). Ce n'est donc qu'une fonction de recherche Google, sans réelle adaptation possible à un site "hôte". Deux options sont cependant possibles : avec ou sans filtre parental.

- Offre "**Free WebSearch + SiteSearch**" (<http://www.google.com/services/free.html>). Cette offre est déjà plus intéressante, ou en tout cas plus personnalisable, que la précédente. Elle permet d'effectuer à la fois des recherches sur le Web ("WebSearch") ET des recherches internes sur le site web sur lequel le formulaire est installé ("SiteSearch"). D'autre part, il est possible d'adapter en partie la page de résultats à vos couleurs et à la charte graphique de votre site. Les champs qu'il est possible de paramétrer sont les suivants :
 - Couleur de fond de la boîte de recherche "Google"
 - Insertion du logo du site "client"
 - Couleur ou image de fond
 - Couleur des liens
 - Couleur du texte.

L'interface de personnalisation, disponible en ligne dès votre inscription, est la suivante :

Customization Settings

SiteSearch Option

Enter domain(s) (ex. google.com): Separate multiple domains with a semi-colon.
 There are approximately **1090** pages for **abondance.com** in our index.

Optional Customization of Results Page

URL (location) of your logo:
ex: http://www.yoursite.com/images/logo.gif

Size of your logo (in pixels): Width: Height:

Alignment of your logo:

Colors can be entered in either text or hexadecimal form, (e.g. black or #000000).

Google logo background color:

Background (BGCOLOR) color:

Text (TEXT) color:

Link (LINK) color:

Visited link (VLINK) color:

Active link (ALINK) color:

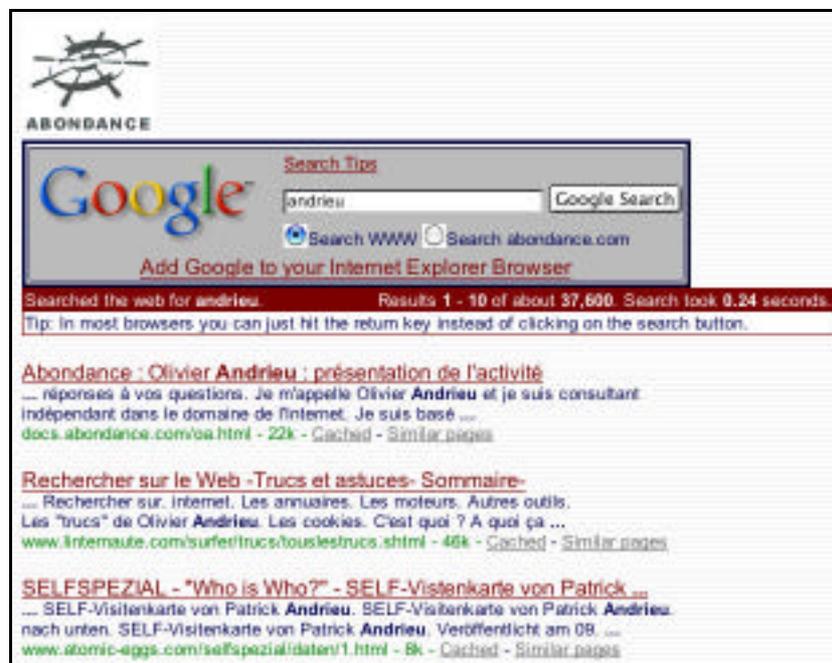
Background URL:

Alternate text color:

Important text color:

Faint text color:

En prenant en compte ces possibilités, voici un exemple de page de résultat (absolument excécrable, je suis d'accord, mais il s'agissait ici de montrer les possibilités de l'option plutôt que de faire un beau document ; je ne doute pas un instant que vous êtes capables de faire mieux ;-)) proposée au travers de cette fonctionnalité :



La boîte de dialogue propose deux possibilités : une recherche globale sur le Web ou spécifique sur le site en question, Abondance, dans notre exemple. Attention : l'outil n'effectue pas une recherche sur TOUTES les pages du site Abondance, mais uniquement sur celles déjà indexées par Google et donc "trouvables" sur le moteur de recherche "classique".

Ainsi, l'option "SiteSearch" de Google effectue ses recherches, dans notre exemple, sur 1090 pages web du site, ce qui est déjà beaucoup mais pas totalement exhaustif .

Ceci dit, le pourcentage de pages indexées par Google dans son index par rapport au nombre de pages total du site est quand même excellent, mais la recherche, quoi qu'il arrive, ne sera pas obligatoirement exhaustive. D'autre part, d'un site à l'autre, ce pourcentage peut fortement changer. Google n'effectue pas une indexation spécifique et exhaustive de votre site si vous souscrivez à l'offre "Free SiteSearch". Autant le savoir...

Bref, il s'agit ici d'une offre intéressante parce que gratuite, mais encore limitée dans les possibilités de "customisation" des pages de résultats.

- **Google API** (<http://www.google.com/apis/>). Nous en avons parlé il y a quelques temps (voir lettre R&R de juillet-août 2002), Google propose depuis quelques mois le service "Google Web API". L'initiative est intéressante et assez innovante dans le domaine des moteurs de recherche : il s'agit de mettre à la disposition des développeurs informatiques une bibliothèque de fonctionnalités (API signifie "Application Programming Interface") utilisables, comme un "Web Service", et qui permettent d'utiliser l'index de Google pour bâtir de nouveaux programmes autour de la recherche d'information sur le Web. Il est techniquement possible d'écrire des applicatifs Java, Perl, PHP, C++ ou grâce à n'importe quel langage qui supporte les "Web Services". Cette bibliothèque de fonctionnalités permet de créer un moteur de recherche à l'image de votre site... si vous avez des talents de programmeur ! Car rien n'est "packagé" dans cette offre (ce n'est pas son but). D'autre part, et ce n'est pas là le moindre des inconvénients, l'utilisation des "Google API" est limitée à 1 000 interrogations de l'index de Google chaque jour. Très limité si vous avez quelques ambitions sur le Web ! Enfin, l'index utilisé par les Web API est un sous-ensemble de celui utilisé "en réel" par le moteur de recherche. Fraîcheur de l'information et exhaustivité non garanties, donc.

- **University Search** (<http://services.google.com/googleuniv/login>). Google propose aux universités la possibilité de personnaliser un moteur de recherche de façon assez partielle mais intéressante. Dans un premier temps, il est possible, un peu comme dans l'option "Free WebSearch", de personnaliser certains champs de la page de résultats :

Your Google's University Search Customizations

Use this form to edit and preview your customizations.

Customization of Results Page

Colors can be entered in either text or hexadecimal form, (e.g. black or #000000).

Google logo background color:

Alternate text color (default green):

Alternate text color is used to present the URLs and error messages.

Faint text color (default light gray - #666666):

Faint text color is used to present cached/similar pages links and the number of results.

Output Language:

Optional WebSearch with SiteSearch: (uncheck box to have just SiteSearch)

Puis, il est possible de rajouter un en-tête aux pages de résultats sous la forme de code HTML :

Provide us header HTML so your search results pages will match the look & feel of your site:

```
<!-- HTML Header START -->
<html>
<head><title>Cobranded Google University Search - Sample HTML</title>
<style><!--
body {font-family: arial,sans-serif}
div.nav {margin-top: 1ex;}
div.nav A,span.nav {font-size: 10pt;font-family: arial,sans-serif}
div.nav A,span.big {font-size: 12pt;color: #0000cc}
div.nav A {font-size: 10pt;color: black;}
--></style>
</head>
<body bgcolor=#FFFFFF text=#000000 link=#0000cc
vlink=#000000 alink=#000099>

<H1>Exemple de recherche sur l'université de Strasbourg</H1>

ici le contenu de l'en-tête

<!-- HTML Header END -->
```

Ainsi qu'un bas de page :

Provide us footer HTML so your search results pages will match the look & feel of your site:

```
<!-- HTML Footer START -->
Google University Search Customers have complete
control over the header and footer html.
<br>
Create the Footer HTML to fit the look & feel of your site.

<p><center><font size=-1>
Copyright ©copy,2000 Google Inc. -
<A HREF=#about.html>About</A> -
<A HREF=#help.html>Help</A>
</font></center>

</body>
</html>

<!-- HTML Footer END -->
```

A partir de ces données, l'applicatif proposera une page de résultat légèrement personnalisée (en prenant les informations du premier formulaire), avec, en haut de page, l'en-tête proposée :



Et en bas de page, le texte convenu à l'avance dans l'interface d'administration :



Là encore, il s'agit d'une solution intéressante (strictement réservée aux universités cependant), mais non satisfaisante encore pour une utilisation professionnelle (même si certaines astuces

peuvent permettre d'obtenir un "look" des pages plus conforme à la charte graphique d'un site, voir ci-après)...

Comparatif des solutions gratuites Google :

	POUR	CONTRE
Buddy link to Google	<ul style="list-style-type: none"> - Gratuit - Très simple - Très rapide d'installation - Recherche sur l'index complet de Google 	<ul style="list-style-type: none"> - Pas de personnalisation possible
Free WebSearch / SiteSearch	<ul style="list-style-type: none"> - Gratuit - Très rapide (personnalisation effectuée en ligne) - Possibilité d'insérer un logo et de légèrement personnaliser la page de résultats - Possibilité de "mixer" une recherche sur le Web avec une recherche sur le site 	<ul style="list-style-type: none"> - Peu de personnalisation possible - Recherche "Web" uniquement sur une partie de l'index total de Google - Pas d'indexation spécifique du site web (toutes les pages ne sont pas indexées) pour l'option "SiteSearch".
Google API	<ul style="list-style-type: none"> - Gratuit - Possibilité assez complètes 	<ul style="list-style-type: none"> - Nécessité de développements spécifiques en langage de programmation - Limitation à 1 000 interrogations/jour - Utilisation d'un sous ensemble de l'index réel pour les recherches web
UniversitySearch	<ul style="list-style-type: none"> - Gratuit - Possibilité d'indiquer un en-tête et un pied de page entièrement personnalisé - Pas de publicités Google - Possibilité de "mixer" une recherche sur le Web avec une recherche sur le site - Recherches sur l'index complet de Google. - Informations statistiques sur l'utilisation du site. 	<ul style="list-style-type: none"> - "Look" des réponses peu personnalisable - Accessible uniquement aux universités.

Si vos ambitions sont supérieures à ce que peuvent proposer les solutions gratuites de Google (et il y a fort à parier qu'elles le seront ou qu'elles le deviendront rapidement), vous devrez donc vous tourner assez vite vers les solutions payantes...

Installations payantes

Google propose, bien sûr, de nombreuses solutions payantes qui vous permettent d'insérer son moteur sur vos pages.

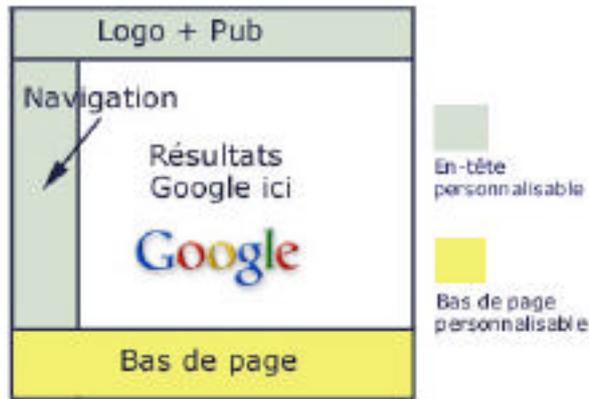
- **Silver Search** (http://www.google.com/services/silver_gold.html)

Ce service coûte 599 \$US (609 euros, environ 4 000 FF) par mois, plus 10 \$US (10,17 euros, environ 67 FF) par millier de requêtes au-delà du premier million.

L'offre propose, comme la prestation disponible gratuitement pour les universités (voir ci-dessus), un en-tête et un bas de page paramétrables, la possibilité de rajouter des publicités qui vous sont propres, un rapport mensuel sur les recherches effectuées sur votre site, une personnalisation selon tel ou tel langage et une possibilité de prendre en compte une recherche "SiteSearch" sur votre site web. Les publicités Google ("Premium Sponsorship" au CPM et "AdWords Select" au CPC)

ne sont pas affichées sur les pages de résultats.

Bref, grosso modo, il s'agit de l'offre proposée gratuitement aux universités, mais de façon payante pour les entreprises. L'intérêt est qu'il est possible de personnaliser l'en-tête et le bas de page des listes de résultats, mais le jeu en vaut-il la chandelle, notamment financièrement, par rapport à l'offre "Site Search" ? A voir ! Car tout le milieu de page (les résultats) reste totalement "Google-like" et non personnalisable ! Notons cependant (et c'est important) qu'en jouant bien sur les en-tête et pied de page, il est possible d'insérer les liens "Google" dans une cellule d'un tableau, ce qui permet d'obtenir un résultat assez proche d'une page de résultats proposant une charte graphique "classique" :



- **Gold Search** (http://www.google.com/services/silver_gold.html)

Ce service coûte, pour sa part, 1 999 \$US (2033 euros, soit environ 13 340 FF) par mois, plus 8 US\$ (8,14 euros, soit plus de 53 FF) par millier de requêtes au-delà des 4 premiers millions.

Les particularités de l'offre sont identiques à celles indiquées ci-dessus (Silver Search).

Il s'agit d'une offre similaire à "Silver Search", mais dont la tarification a été adaptée à des sites proposant un plus fort trafic.

- **Custom WebSearch** (<http://www.google.com/services/custom.html>)

Il s'agit de la "Rolls" des possibilités d'intégration de Google à un site. Dans ce cas, tout est paramétrable et modifiable. Les résultats sont proposés par Google en format XML ou selon un protocole propriétaire. Le contrôle du rendu des pages de résultats est complet.

Des informations statistiques très précises (trafic, nombre de requêtes effectuées, mots clés le plus souvent demandés, etc.) sont également proposées par le moteur.

Plusieurs options sont disponibles : filtre familial par défaut ou non, recherche dans plusieurs langages, etc.

Un bon exemple de ce type d'offre peut être visualisé sur le site <http://www.virgin.net/> (option "World" du formulaire de recherche), ou, en France, sur le site de Free (<http://www.free.fr/>).

Evidemment, le site le plus connu utilisant cette possibilité est Yahoo!.

Mais avoir à sa disposition autant de flexibilité et de possibilité disponibles n'est pas gratuit. Pour prendre en compte l'offre "Custom Web Search" sur votre site, il vous faudra vous acquitter de l'achat d'une licence annuelle (50 000 US\$ au moins). Un coût dépendant du nombre de requêtes semble également être demandé. Nous vous fournirons l'information dès que nous obtiendrons plus d'information sur ce chiffre que nous n'avons pas pu obtenir encore. A noter que les informations tarifaires pour cette offre sont données à titre purement indicatif. Nous devrions recevoir très bientôt d'autres informations pour recouper celles déjà en notre possession. Nous vous en ferons part dès réception.

- **Custom SiteSearch** (<http://www.google.com/services/customsitesearch.html>)

Il s'agit là du pendant de l'offre "Custom WebSearch", mais spécifiquement adapté à la recherche à l'intérieur d'un site web, sur ses pages internes. Tout ce qui a été dit sur les possibilités "Custom WebSearch" reste valable, si ce n'est que l'index utilisé est bien entendu spécifique et comprend TOUTES les pages de votre site. Contrairement, donc, à l'offre "Silver/Gold", un "crawl" spécial (quotidien ou hebdomadaire) de vos pages est effectué spécifiquement par Google. L'index utilisé est donc exhaustif et à jour.

L'option "Key match" est également intéressante : pour certains mots clés, vous pouvez indiquer "en dur" quelles urls peuvent être affichées dans les premières positions. Dans ce cas, l'algorithme de pertinence de Google ne sera pas utilisé.

Mais, là aussi, l'offre n'est pas gratuite. En coût annuel, il faudra compter de l'ordre de 195 000 US\$. Un coût dépendant du nombre de requêtes est également demandé. Nous vous fournirons l'information dès que nous obtiendrons plus d'information sur ce chiffre que nous n'avons pas pu obtenir encore. A noter que les informations tarifaires pour cette offre sont données à titre purement indicatif. Nous devrions recevoir très bientôt d'autres informations pour recouper celles déjà en notre possession. Nous vous en ferons part dès réception.

• **Search Appliance** (<http://www.google.com/appliance/>)

Les "Search Appliance" se présentent sous la forme de "boîtes noires" permettant d'effectuer des recherches spécifiquement sur les intranets d'entreprises. Il en existe un gamme variée, depuis la simple "boîte à pizza" jusqu'aux armoires de rack pour les gros sites. Cette offre n'étant pas disponible en Europe (elle devrait le devenir en 2003) et étant plutôt réservée à des intranets, nous ne la développerons pas ici.

Comparatif des solutions payantes Google :

	Coût mensuel	Coût à la requête	POUR	CONTRE
Silver Search	599 US\$	10 US\$ par millier de requêtes au-delà du premier million.	<ul style="list-style-type: none"> - Possibilité d'indiquer un en-tête et un pied de page entièrement personnalisés. - Pas de publicités Google. - Possibilité de "mixer" une recherche sur le Web ("WebSearch") avec une recherche sur le site ("SiteSearch"). - Informations statistiques sur l'utilisation du moteur. - Interface disponible dans de nombreuses langues. 	<ul style="list-style-type: none"> - Assez cher - "Look" des liens de réponses non personnalisable - Pas d'indexation spécifique du site web pour la fonction "SiteSearch" (toutes les pages ne sont pas obligatoirement indexées, seules celles disponibles dans l'index du moteur sont "trouvables"). - La recherche "SiteSearch" ets indissociable de la recherche "WebSearch". - Recherches effectuées sur une partie de l'index complet de Google.
Gold Search	1 199 US\$	8 US\$ par millier de requêtes au-delà des 4 premiers millions.	Idem que Silver Search	Idem que Silver Search
Custom Web Search	50 000 US\$	N.C.	<ul style="list-style-type: none"> - Page de résultats totalement paramétrable. - Informations statistiques mensuelles très complètes. - Contrôle complet du service. - Recherches sur l'index complet de Google. 	<ul style="list-style-type: none"> - Cher (réservé aux grands sites portails).
Custom	195 000 US\$	N.C.	- Page de résultats	- Cher (réservé aux grands

<p>SiteSearch</p>			<p>totalemment paramétrable. - Informations statistiques mensuelles. - Contrôle complet du service. - Index de pages réalisé spécifiquement : exhaustif et complet.</p>	<p>sites portails).</p>
--------------------------	--	--	--	-------------------------

Conclusion

En conclusion, on peut dire que Google propose de très nombreuses solutions d'intégration de son moteur sur un site web distant. Son offre et sa gamme de prestations est certainement l'une des plus larges à l'heure actuelle. On peut cependant regretter que la solution de loin la plus intéressante (Custom WebSearch) soit très largement hors de portée, financièrement parlant, d'un site "moyen". Peut-être qu'un coût annuel, un coût d'entrée, moins important, mais avec des redevances plutôt indexées sur le nombre de requêtes et une offre commerciale proposant, par exemple, de partager les recettes publicitaires des "Premium Sponsorship" et "Adwords Select" aurait été mieux acceptées et appréciées des sites ayant peu de moyens, tout en augmentant la visibilité du moteur sur le Web. Mais peut-être n'est-ce que partie remise...