

## Exorbyte propose des Google Suggests "intelligents"

[Retour au sommaire de la lettre](#)

*Vous connaissez peut-être l'application Google Suggest, qui affiche des propositions de mots clés au fur et à mesure de la saisie d'une requête dans le formulaire de recherche du moteur. La société allemande Exorbyte propose depuis peu une nouvelle génération de ce type d'outils, très utile notamment sur des sites de commerce électronique ou permettant d'effectuer des recherches dans des bases de données proposant de nombreuses informations. Explications...*

La plupart d'entre vous connaissent certainement les "Google Suggest" (<http://www.google.com/webhp?complete=1&hl=en>), applicatif qui propose des mots clés au fur et à mesure de la saisie de la requête :



Web	Images	Groupes	Annuaire	Actualités
référencement				
référencement				1,780,000 results
référencement gratuit				635,000 results
référencement site				104,000 results
référencement google				802,000 results
référencement manuel				189,000 results
référencement moteurs recherche geneve				649 results
référencement international				276,000 results
référencement québec				121,000 results
référencement de site				1,780,000 results
référencement sur mesure				257,000 results

Sur la lancée de cet outil, de nombreux sites (dont le moteur [Mozbot.fr](http://Mozbot.fr)) ont proposé des services similaires. On peut également noter l'outil Google Synonymes (<http://david.dotgeek.org/google/syn/>) qui, bien qu'il semble un peu "bricolé" et qu'il ne fonctionne pas sur toutes les plates-formes, propose un ensemble de synonymes du mot clé en cours de saisie.



Il existe de très nombreux outils de ce type (voir <http://ressources.abondance.com/generateur-mot-cle.html>), souvent très semblables dans leur forme.



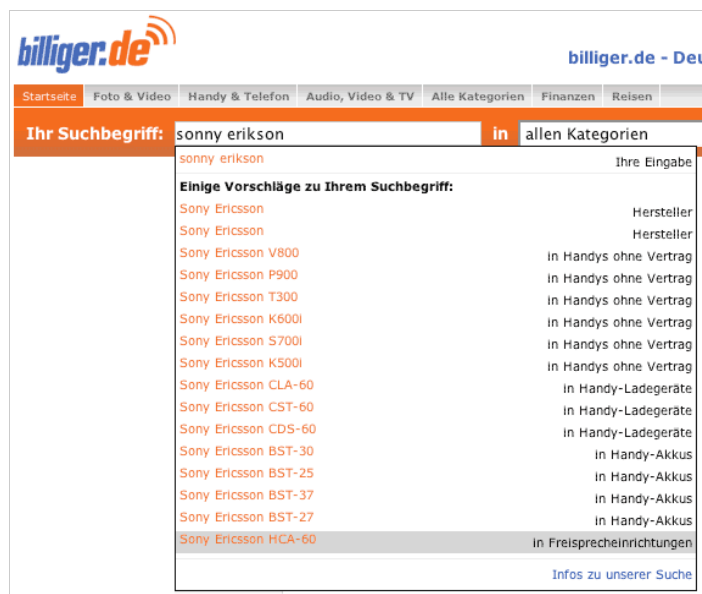
Mais une société d'origine allemande, Exorbyte (<http://www.exorbyte.com/>), a décidé d'aller plus loin dans ce domaine et de proposer une offre de "fenêtres contextuelles d'aide à la saisie", baptisée "MatchMaker Suggest", intégrant différentes informations complémentaires pour aider l'internaute à mieux formuler sa question.

Voici quelques exemples de ce que propose Exorbyte :

- Sur le site Billiger (<http://www.billiger.de/>), site de commerce en ligne, l'outil propose des modèles ainsi que la catégorie dans laquelle il est possible de trouver l'article :



L'idée est de joindre à l'outil un système de correction d'orthographe. Exemple avec la saisie de "sonny erikson" où le système reconnaît la marque scandinave :



L'internaute peut alors, en un clic, identifier le modèle (ici les téléphones portables de marque Sony Ericsson) sans connaître parfaitement l'orthographe de la marque désirée. Il a également accès rapidement à la liste des modèles disponibles tout en restant dans le formulaire de recherche.

- Sur le site JobaNova (<http://www.jobanova.de/>) de recherche d'emploi, le système est similaire si l'on tape un intitulé d'emploi ou un nom de ville :



Bitte nur Berufs- und/oder Ortsbezeichnungen verwenden

konst		SUCHEN
Raumpfleger/in	7846x	Konstanz (3)
Elektromechaniker/in	7846x	Konstanz (2)
Reinigungskraft	78467	Konstanz (1)
LKW-Fahrer/in	78467	Konstanz (1)
Medienkünstler/in	78467	Konstanz (1)
Mediendesigner/in	78467	Konstanz (1)
Klempner/in	78467	Konstanz (1)
Assistent/in	78467	Konstanz (1)
Dipl.-Ing. Informatik	78467	Konstanz (1)
Plattenleger/in	78467	Konstanz (1)
Mosaikleger/in	78467	Konstanz (1)
Fliesenleger/in	78467	Konstanz (1)
Assistent/in	78467	Konstanz (1)
Werkzeugmacher/in	78467	Konstanz (1)
Pizzabäcker/in	78467	Konstanz (1)
Schweißer/in	78467	Konstanz (1)
Maschinenbediener/in	78467	Konstanz (1)
Lagerarbeiter/in	78467	Konstanz (1)
Engineer/in	78467	Konstanz (1)
Produktmanager/in	78467	Konstanz (1)
WIG-Schweißer/in	78467	Konstanz (1)

... tippen Sie einfach weiter um exaktere Ergebnisse zu bekommen ...

Là encore, plusieurs informations sont affichées dans la fenêtre : poste à pourvoir, nombre d'annonces, code postal, ville. Il est alors possible d'aller beaucoup plus rapidement à l'information.

Idem si on saisit le nom d'un poste à pourvoir :

Bitte nur Berufs- und/oder Ortsbezeichnungen verwenden

platten		SUCHEN
Plattenleger/in	84424	Isen (1)
Plattenleger/in	25917	Leck (1)
Plattenleger/in	99831	Ifta (1)
Plattenleger/in	66706	Perl (1)
Plattenleger/in	46459	Rees (1)
Plattenleger/in	71404	Korb (1)
Plattenleger/in	71732	Tamm (1)
Plattenleger/in	56332	Löf (1)
Plattenleger/in	75365	Calw (1)
Plattenleger/in	38274	Elbe (2)
Plattenleger/in	5942x	Unna (2)
Plattenleger/in	59065	Hamm (1)
Plattenleger/in	07548	Gera (1)
Plattenleger/in	07749	Jena (1)
Plattenleger/in	49757	Lahn (1)
Plattenleger/in	8907x	Ulm (4)
Plattenleger/in	241xx	Kiel (2)
Plattenleger/in	26316	Varel (1)
Plattenleger/in	32257	Bünde (1)
Plattenleger/in	44532	Lünen (1)
Plattenleger/in	94209	Regen (1)

... tippen Sie einfach weiter um exaktere Ergebnisse zu bekommen ...

On a alors accès aux villes où ces postes sont disponibles ainsi qu'au nombre d'annonces en cours...

- Un outil basé sur le site communautaire de questions-réponses Lycos IQ (<http://iq.lycos.fr/>) a également été développé par Exorbyte (<http://www.mmsuggest.de/lycos-fragen/>) :

**Lycos-IQ: Fragen Sie unser Suggest!**

Search for:

Questions:	Warum bekommt man bei Google so schlechte Antworten?	Google
	Wie funktioniert das verteilte Google Rechenzentrum und wo wird die zentrale Intelligenz gehalten?	Google
	Wo krieg ich einen Code bzw. die Einladung fuer google Analytics her?	google
	Wie kann man bei Google nur nach bestimmten Dokumenttypen suchen?	Google
	Wie kann ich Google sofort auf meine Homepage locken?	Google
	Wie bekomme ich einen Gmail -Account bzw. eine Einladung, diesen Maildienst zu nutzen?	email,google
	Wer schickt mir eine Einladung für Google Mail?	Google Mail
	Hat Google viel Macht?	Macht,Google
	Google Earth	Google Earth,Strassen,suchen
	Kann ich die Satellitenaufnahme von Google earth auch ohne leistungsfähige Internetverbindung sehen?	google earth
	Google-Earth Extras	Google Earth
	Google earth?	Google Earth,Satellitenbilder,internet

Powered by exorbyte

Vous tapez le début d'une question ou d'un mot clé et l'outil suggère toutes les questions qui sont déjà dans la base de données et qui contiennent ce terme. Là encore, les potentialités de développement de ce type d'outil sont fortes, notamment sur des systèmes d'aide en ligne ou sur des intranets...

La société a également développé des modules d'interrogation de Wikipedia (<http://knecht.cis.uni-muenchen.de/exorbyte/wiki/>) qui propose des mots clés connexes avec la langue dans laquelle l'article est disponible, ainsi qu'une recherche de noms (<http://knecht.cis.uni-muenchen.de/exorbyte/names/>) sur la base de données extraites du Web (avec reconnaissance des orthographes approximatives) :

**exorbyte**  
Approximate Search Solutions

C. Guenthner
L. Guenthner
Joe Guenthner
Ken Guenthner
Rob Guenthner
Greg Guenthner
Paul Guenthner
Alvin Guenthner
David Guenthner
Diane Guenthner
Franz Guenthner
Scott Guenthner
Vince Guenthner
Alfred Guenthner
Edmund Guenthner

Suggestions provided by exorbyte

L'idée de base du système d'Exorbyte est de fournir non seulement une liste de termes "commençant par" ce qui est saisi dans le formulaire de recherche du moteur, mais surtout d'y ajouter des données permettant à l'internaute de mieux faire son choix...

L'outil "MatchMaker Suggest" a été réalisé en Ajax et l'applicatif réside en RAM, allant chercher ses informations dans un fichier texte qui est retravaillé par le système (le traitement d'un fichier de 30 millions d'entrées prenant environ 30 minutes). L'application fonctionne sous Windows, Solaris et Linux. Des APIs sont disponibles pour les langages suivants : Sockets, C++, COM, PHP, Java, Tcl, Python. Des extensions OCR (reconnaissance optique de caractères) sont également disponibles.

Exorbyte n'est peut-être pas si inconnue que cela si vous êtes un lecteur assidu de la lettre "Recherche & Référencement". En effet, la société a été créée par Benno Nieswand et Franz Guentner, "Professor for Computational Linguistics" à l'université de Munich, qui nous avait reçu en juillet 2002 afin de nous expliquer la nature de ses travaux pour le moteur de recherche Fast / AllTheWeb (<http://abonnes.abondance.com/archives/2002-07/index.html>).



En marge de sa collaboration avec Fast, il a également travaillé sur le moteur Accoona (<http://www.accoona.eu/>) et notamment sur sa fonction "Super Target your Search", qui permet d'obtenir des informations connexes au sujet d'une requête sur le Web ou l'actualité.

Franz Guentner a bien voulu répondre à cinq questions sur ce nouvel outil...

**- Franz, bonjour et merci de répondre à nos questions. Quel est le principal avantage de MatchMaker Suggest par rapport aux systèmes de suggestions actuels (Google Suggest, Surfswax, etc.) ?**

A la très grande différence des "Suggests à la Google et autres" qui ne font que "lire" une liste de mots de gauche à droite, notre technologie consiste à chercher "intelligemment" dans les données d'un index. Nous trouvons donc toujours la "meilleure" réponse même si la requête contient des erreurs (ce qui, sur un moteur traditionnel, échouerait tout de suite). Si vous faites une toute petite faute sur Google Suggest ou si vous employez une variante morphologique, vous ne trouvez plus rien. En d'autres termes, nous "cherchons" approximativement dans les données disponibles. Matchmaker est basé sur des algorithmes et des implantations propriétaires d'une vingtaine d'algorithmes relevant de la recherche approximative sur des chaînes. En gros nous travaillons sur des bases de données énormes à peu près 100 fois plus rapidement que pour les approches courantes basées sur la programmation dynamique. Nous constatons des temps de réponse de moins de 10 ms même quand il y a des dizaines de millions de données dans l'index. Dans certaines applications nous pouvons aussi avoir recours à une présentation structurée des résultats comme par exemple chez <http://www.cooceo.de/>.

**- Que faut-il ajouter à un site web existant pour intégrer votre solution de recherche ? Quelle est la procédure technique ?**

On peut soit utiliser Matchmaker comme un système complet où on peut indexer un nombre quelconque de champs (un peu comme dans une base de données traditionnelle). Ou on peut aussi se servir uniquement de la fonction "Suggest" sur les champs principaux de l'application. Cette dernière méthode ne requiert aucun changement dans la plate-forme déjà installée et elle peut être réalisée en un jour ou deux (dans beaucoup de cas, l'index utilisé réside même chez nous et non pas chez le site qui s'en sert..).

**- Combien coûte votre solution ? Avez-vous une échelle de prix à nous suggérer ?**

Cela dépend si l'on licencie le système complet ou juste la fonctionnalité "suggest". Et dans les deux cas le prix est fonction de la taille de la base de données et du nombre de requêtes. On commence typiquement avec un coût de 30K - 50K euros sur le système complet et à quelques centaines d'euros par mois pour le "Suggest" seul. Allez sur <http://mmsuggest.de/> pour voir comment il est facile de se servir de notre technologie...

***- Avez-vous pu mesurer les avantages d'une solution comme la vôtre pour l'internaute qui l'utilise et le site web qui l'implémente ?***

Nous avons constaté que nos clients dans l'e-commerce, où la réussite se mesure en nombre de transactions conclues, augmentent typiquement leur revenu par 10% à 20% très peu de temps après avoir installé notre technologie. Et ce n'est pas surprenant puisqu'il n'y a plus d'échecs causés par la forme de la requête et - qui plus est - l'utilisateur voit toute suite quels sont les autres objets qui pourraient l'intéresser. Nous annonçons - sans rire - qu'on cherche 100 fois plus vite avec notre recherche qu'avec la recherche traditionnelle.

***- Envisagez-vous des installations en langues anglaise et française ? L'outil est-il prêt pour ses langues ?***

Les fonctionnalités de base de Matchmaker sont complètement indépendantes des langues en questions. On tourne d'ailleurs aussi avec un codage Unicode/utf8. Mais nous pouvons évidemment ajouter des raffinements linguistiques propres à une langue donnée comme par exemple une forme de résolution phonétique ("soundex"), la normalisation morphologique, des tables de synonymes et beaucoup d'autres outils linguistiques. C est après tout notre spécialité...

***Merci Franz.***