

# Exalead

## *en profondeur*

<http://www.exalead.fr>



**Jean-Pierre LARDY**

UNIVERSITE CLAUDE BERNARD - URFIST de LYON

courriel : [lardyATuniv-lyon1.fr](mailto:lardyATuniv-lyon1.fr) serveur : <http://urfist.univ-lyon1.fr>

*maj : sept 2008*

## 1 - Présentation

**Exalead** est un moteur de recherche généraliste sur le web mondial, conçu en France.

La société **Exalead** a été fondée en 2000 par Patrice Bertin et François Bourdoncle<sup>1</sup>. Cet ingénieur de l'Ecole des Mines de Paris avait participé à la création du moteur de recherche AltaVista, vedette des moteurs de recherche dans la seconde moitié des années 90 (aujourd'hui technologie propriété de Yahoo). Il y avait développé la fonction *Refine* qui permettait d'affiner une recherche en proposant plusieurs mots clés proches par analyse statistique du contenu des pages trouvées par la requête.

En octobre 2006, **Exalead** a lancé officiellement une version grand public de son moteur. Le modèle économique repose sur les liens sponsorisés externalisés à Yahoo ! .

*La technologie d'Exalead est basée uniquement sur des **algorithmes statistiques** qui peuvent traiter des corpus très importants. Ils permettent de proposer des fonctions « intelligentes » que les systèmes linguistiques ne peuvent fournir comme le repérage automatique des expressions, les fautes d'orthographe. La lemmatisation permet ainsi de retrouver les singuliers/pluriels mêmes irréguliers. En effet la nature hétérogène, gigantesque et multilingue du web mondial interdit dans la pratique les traitements linguistiques.*

*Pour proposer des fonctionnalités avancées de recherche sémantique multilingue et de cartographie des résultats de recherche, Exalead a signé un partenariat technique et commercial avec Lingway, éditeur de Lingway KM, moteur de text mining.*

**Domaines couverts** : multidisciplinaire

**Langue** : multilingue

**Volume** : en octobre 2006 **Exalead** possédait une base de plus de 8 milliards de pages et plus de 2 milliards d'images indexées (équivalent à Google)

**Mise à jour** : mensuelle

**Formats de fichier** : HTML, Word, Excel, PowerPoint, Rich Text Format, WordPerfect PDF, TXT, Flash

---

<sup>1</sup> Voir interview de François Bourdoncle dans "**Recherche et Référencement**", mai 2002, **Exalead : une technologie d'avance**

## 2 - Interrogation du web

<http://www.exalead.fr> Exalead propose son interface dans plusieurs langues.

Un lien *Préférences* permet :

- de choisir la langue des documents recherchés (toute langue par défaut),
- d'activer le filtrage du contenu adulte,
- de choisir le format d'affichage des résultats avec ou sans vignette de prévisualisation en particulier.

Une des caractéristiques d'Exalead consiste à combiner les pages récupérées par son robot et les sites de l'annuaire coopératif Open Directory<sup>2</sup>. C'est un moyen de lever les ambiguïtés du langage naturel et une aide à la formulation des requêtes.

Exemple: soit une recherche sur "chirurgie laser", les résultats proposent les catégories "Santé" ou "Economie Commerce". On peut ensuite choisir une catégorie pour affiner sa recherche.

### Recherche simple

Quatre onglets de recherche sont proposés : par défaut le web, puis la recherche d'images, la recherche dans l'encyclopédie Wikipedia et la recherche de video.

**Dans le web**, on peut rechercher soit des pages sur le web multilingue, des pages en français ou hébergées sur des serveurs en France.

Le formulaire utilise la syntaxe maintenant standard des moteurs de recherche : ET implicite, recherche d'expression avec les guillemets plus des opérateurs permettant de préciser ses recherches : **proximité** : opérateur NEAR , la **troncature** : astérisque (\*) correspond à la troncature droite et illimitée, et les **parenthèses**.

exemple : seul Exalead permet de composer une telle équation de recherche de ce type :

((développement\* NEAR durable\*) OR (commerce\* NEAR équitable\*))

Pour affiner une recherche, Exalead ouvre une fenêtre, à droite des résultats, et propose différents filtres selon les résultats obtenus : termes associés, langues, zone géographique, type de fichiers, type de sites (blog, forum), multimedia (Audio, Vidéo, RSS).

En cliquant sur un ou plusieurs critères, on précise la requête. Mais il est toujours possible de revenir en arrière grâce au lien supprimer.

Le bouton Plus de choix augmente les listes de critères et précise les pourcentages de résultats pour les langues, types de documents et situation géographique.

Des liens sponsorisés peuvent être affichés en fonction des questions : ils s'affichent en tête de liste sur fonds bleu clair.

<sup>2</sup><http://dmoz.org>

**Affiner la recherche**

**Vos affinages**

**Annuaire**

📁 Santé [supprimer](#)

**Termes associés**

- [Chirurgie esthétique](#)
- [Epilation laser](#)
- [Greffe de cheveux](#)

**Multimedia**

📡 [RSS](#)

**Langues**

- [Français](#)
- [Néerlandais](#)

**Annuaire**

- 📁 [Médecine](#)
- 📁 [Spécialités médicales](#)
- 📁 [Hôpitaux et cliniques](#)

**Type de fichiers**

- [Acrobat \(.pdf\)](#)

[Plus de choix »](#)



### Recherche avancée

Le lien Recherche Avancée ouvre une fenêtre avec un grand nombre de fonctionnalités absentes des grands moteurs concurrents, indispensable pour effectuer des recherches un peu approfondies sur les web.

Ainsi l'équation suivante : `spelllike:lekemia country:dz language:ar` permet de retrouver des documents en arabe sur la leucémie publiés en Algérie.

### Recherche d'images



On affiche les informations de chaque image en faisant passer le curseur de la souris dessus.

Les résultats de la **recherche d'images** s'accompagnent d'un lien Recherche Avancée permettant de limiter les résultats par taille, couleurs, catégorie de photos (portrait/paysage) et surtout aux visages. Cette dernière fonctionnalité s'appuie sur la technologie de la société **LTU technologies**<sup>3</sup> issue de travaux de recherche de l'INRIA.

Un filtrage des contenus adulte est modifiable par l'onglet Préférences.

Il est possible de sauvegarder les images sur son disque dur ou en ligne (avec un compte) avec Images Préférées.

### Recherche dans Wikipedia

La **recherche dans Wikipedia** apporte une véritable valeur ajoutée à l'encyclopédie et remplace avantageusement son moteur interne. Suite à une requête, Exalead liste les résultats et propose de préciser sa question grâce à un nuage de mots-clés (tags) colorés selon qu'il s'agit de catégories de Wikipédia, de termes associés, de personnes, d'organisations ou de lieux.

Des drapeaux donnent accès aux résultats des différentes versions linguistiques de Wikipedia.

### Recherche de video

Le moteur recherche sur les principales plateformes de partage de vidéos : le français DailyMotion, Kewego, Metacafe, iFilm, Vodeo, YouTube. Là aussi on peut affiner les résultats en choisissant un site de vidéos en particulier ou une durée : courte (< 1 minute), moyenne (1-15 minutes) ou longue (> 15 minutes).

Un nuage de mots-clés permet de préciser sa question par rubrique ou mot-clé.

<sup>3</sup> <http://www.ltutech.com/fr/>

**Préciser la recherche**



Africa Asia

Biological pest control agents China

**Claire Villemant** Europe

Hyménoptère (nom scientifique) INRA

Italy Japan **Jean Haxaire**

**Jean-Claude Streito**

**Jean-Philippe**

**Tamisier Jean-Pierre**

**Bouquet**

**Lot-et-Garonne** Med J Aust

North America Papua New Guinea

**Pest Sicily Société**

**Entomologique de**

**France** The European **Vespa**

**Vespidé Vespoidea**

 Catégories
  Termes associés

 Personnes
  Lieux

 Organisations

### 3 – Résultats

Chaque résultat a un lien [Ajouter aux raccourcis](#) permettant d'enregistrer l'adresse dans ses favoris Exalead et une vignette de prévisualisation de la page.

Enfin il est possible de s'enregistrer sur le site Exalead (S'identifier). Avec ce compte on peut sauvegarder ses raccourcis et ses préférences et les utiliser à partir de n'importe quel ordinateur.

### 4 – Compléments

**Exalead Toolbar** pour Firefox permet de bénéficier instantanément du moteur de recherche depuis son navigateur.

**Exalead a décliné sa technologie en plusieurs produits :**

- **exalead one:desktop** : recherche sur le poste de travail et le web ;
- **exalead one:workgroup** : recherche sur les disques durs et les réseaux locaux ;
- **exalead one:enterprise** pour les bases de données, les serveurs d'entreprise, le web et les postes de travail ;
- **exalead one:websearch** : le moteur de recherche généraliste sur le web ;
- **exalead one:datacenter** pour le traitement de milliards de documents avec des contraintes de trafic très élevées.

Voir <http://corporate.exalead.com/enterprise/l=fr>

### 5 – Nouveautés

juin 2007 - Exalead intégrera les données scientifiques de l'INIST : 13 millions d'articles scientifiques seront bientôt intégrés dans la base du moteur de recherche d'Exalead grâce à l'accord passé avec l'Institut de l'Information Scientifique et Technique du CNRS (INIST).