

LES OUTILS DE RECHERCHE SUR INTERNET : PISTE POUR UNE ANALYSE THE SEARCH TOOLS ON THE INTERNET : TRACK FOR AN ANALYSIS

François DIERICK
f_dierick@hotmail.com

Cet article offre une grille d'analyse permettant d'étudier sous différents aspects les outils de recherche de l'Internet, à savoir les annuaires et les moteurs de recherche. A partir de cette grille, il est possible de créer des tableaux de synthèse, comparant différents outils. Des exemples de tableaux, complétés pour huit outils de recherche belges et six outils de recherche mondiaux, sont présentés.

This article offers a grid of analysis to enable the study of the Internet search tools such as directories and search engines. Tables are drawn from that grid which compare the different tools. Examples of these tables are presented, for eight Belgian and six world-wide search tools.

INTRODUCTION

Lorsqu'un nouveau protocole naît sur l'Internet et commence à être abondamment exploité, des outils de recherche appropriés voient le jour. Le web a connu en sept années un développement inouï : le nombre actuel de pages web consultables est estimé entre deux et quatre milliards, auxquelles viennent se rajouter quotidiennement entre un et sept millions de nouvelles pages. Les chercheurs de l'Université de Berkeley aux Etats-Unis estimaient en 2001 le nombre total de documents présents sur l'Internet à 550 milliards, dont un seul milliard était indexé ¹.

¹ CENTRE DE DOCUMENTATION DU PÔLE SCIENTIFIQUE. (page consultée le 14 septembre 2002) *Critères d'évaluation des sources d'information d'Internet* [en ligne]
http://www.uco.fr/services/biblio/cdps/web_chiffres.html

Différents outils existent pour consulter le web. Généralement, nous faisons appel aux annuaires et aux moteurs de recherche.

Les annuaires recensent des sites web, et non les pages web, par l'intermédiaire d'une intervention humaine, et organisent ces sites de façon hiérarchique selon une arborescence thématique, différente pour chaque annuaire. Vu la méthode d'indexation, il est difficile, voire impossible, pour les annuaires de couvrir la totalité du web. Les annuaires privilégient la qualité des ressources, en n'indexant que certains sites. Ils possèdent la faculté d'offrir un aperçu de l'existant sur l'Internet relatif au domaine recherché, ainsi que son importance et ses relations hiérarchiques au sein d'une classification.

Les moteurs de recherche sont des outils automatiques qui recensent et indexent le web, privilégiant la quantité des ressources indexées, ignorant la notion de site en indexant toutes les pages de celui-ci. L'indexation des moteurs de recherche est basée sur le full-text et les métadonnées.

Enfin, de nombreux utilisateurs ont recours aux portails comme point d'accès et de recherche sur l'Internet. Un portail est par définition un point d'entrée sur le Net offrant différents services aux internautes, et dès lors, des services de recherche également. Généralement, les portails qui proposent des recherches sur le Net utilisent la base de données d'outils spécialisés dans la recherche.

Il y a un an, les moteurs de recherche et les annuaires étaient estimés à environ 47.000².

Parmi tous ces outils, certains sont généralistes et indexent - ou plutôt tentent d'indexer - le maximum de ressources, et ce, à l'échelle mondiale. La masse de documents, et le caractère dynamique de ceux-ci, est telle qu'il leur est impossible de tout recenser, et indexer. Il est donc très difficile de garder un outil parfaitement à jour, surtout en ce qui concerne les annuaires, demandant un apport humain lors du traitement des données pour l'indexation et le classement hiérarchique, alors que les moteurs de recherche effectuent une indexation automatique.

Analyse des outils de recherche

En seulement quelques années de vie, le web a vu naître - et disparaître - de nombreux sites consacrés à la recherche, derrière lesquels se trouvent généralement des sociétés dont le but est le profit. Parmi tous ces outils, certains offrent entière satisfaction, mais de nombreux autres ne permettent pas d'effectuer de réelles recherches, en raison de la nature de la base de données et du nombre de documents indexés, la mise à jour irrégulière de l'information, de la difficulté d'utilisation du mode d'interrogation de la base,... Il n'est donc pas toujours aisé de trouver des outils de recherche pertinents. Une analyse préalable doit donc être effectuée, basée sur certains critères plus ou moins rigoureux, objectifs, et parfois subjectifs. Différents critères doivent être pris en compte : la qualité de la base, et le nombre de documents recensés ; la facilité d'interrogation ; le type d'interrogation possible (opérateurs booléens, opérateurs de proximité, gestion des mots vides,...), ainsi que la méthode d'indexation utilisée lors du traitement des documents (indexation manuelle, indexa-

tion automatique, bases de l'indexation,...). Une certaine approche de l'outil doit donc être réalisée, avant de pouvoir réellement juger de ses capacités. Pour ce faire, il est intéressant, et plus facile, de compléter une grille qui aborde différents aspects de l'outil analysé, et qui offre ainsi un compte rendu complet et détaillé de ses ressources, du fonctionnement et de ses caractéristiques.

La grille présentée ci-après est composée de quatre parties :

- **une présentation générale de l'outil de recherche analysé**, prenant uniquement en compte sa description. On y retrouve les différents éléments propres qui le composent et le distinguent des autres outils de recherche ;
- **des données concernant la récolte et l'indexation des documents**, permettant de comprendre le traitement de l'information contenue dans les pages web par l'outil de recherche ;
- **des essais de recherche et l'interprétation des résultats obtenus**, afin de pouvoir tester les capacités de recherche de l'outil, les méthodes de recherche et la présentation de résultats, variant plus ou moins entre les outils de recherche. Des tests sont donc effectués :
 - o *pour les moteurs de recherche*, les recherches portent sur certains termes et différentes combinaisons de ceux-ci, vérifiant ainsi l'éventuelle possibilité de recherches booléennes ou d'autres types de recherche ;
 - o *pour les annuaires*, les recherches portent sur des domaines précis, afin de vérifier si l'utilisateur peut naviguer dans la hiérarchie de l'annuaire de façon innée, et trouver des résultats satisfaisants.
- **l'interface du site**, présentant brièvement le design du site (publicités, lisibilité,...) et son propre signalement dans les autres outils de recherche.

Pour terminer l'analyse de l'outil, une **évaluation** peut également être effectuée, en se basant sur certains critères provenant de la grille, comme par exemple la

² BRIGHT PLANET. (page consultée le 14 septembre 2002). *BrightPlanet - Deep Content* [en ligne]
http://www.brightplanet.com/deepcontent/index_cp.asp

quantité, voire l'absence, des liens cassés ou inexistantes ; la présence d'une aide en ligne pour effectuer de recherche ou soumettre un site ; les différentes possibilités de recherche (opérateurs booléens, de proximité,...),... D'autres critères, plus subjectifs, peuvent également intervenir,

comme la facilité de navigation, la facilité de contacter le webmaster, ...

Cette grille, appliquée telle quelle ou modifiée en fonction des besoins, devra être remplie de façon similaire pour chaque outil analysé.

Grille de recherche

DESCRIPTION

- **Généralités**

Nom

Nom complet du moteur de recherche, annuaire ou portail

Type

Type d'outil : annuaire, moteur de recherche, portail

Nature du site

Le site est-il de nature personnelle ou de nature commerciale ? Le but est-il dans ces conditions d'informer, ou d'offrir une information couverte sous de la publicité ? L'objectif est-il un objectif d'information, un objectif commercial ou promotionnel ?

URL de la page principale

URL de la page d'accueil du site

URL de la page à partir de laquelle une recherche peut être effectuée

URL de la page où l'écran de saisie est présent, ou la page où la navigation dans l'annuaire est possible

URL de soumission de site

URL de la page où l'on peut soumettre un site, ou le modifier éventuellement

Couverture

Le web mondial, le web belge, le web belge francophone,...

Langue(s) du site

Langues dans lesquelles le site est accessible

Date de création

Année

Dernière modification

Date de la dernière modification du site

- **Particularités**

Auteurs et webmasters du site

Le nom et prénom des auteurs et webmasters du site, avec contact e-mail

Nombre d'employés

Nombre de personnes employées et leur formation générale

Données de contact du site

Adresse, téléphone, fax, e-mail

Nombre de sites indexés

Nombre de sites indexés à la date de l'analyse du site

Fréquentation du site

Statistique de fréquentation les plus récentes possibles

Périodicité de mise à jour

La périodicité de mise à jour de la base de données

RECOLTE ET INDEXATION DES DOCUMENTS

- **Documents indexés**

Type de documents indexés

Le web, les adresses e-mail, les MP3, les documents picturaux,...

Langue(s) des documents indexés

Français, anglais, néerlandais, allemand,...

- **Méthode de récolte des documents**

Collecte des documents

Ici est mentionnée la méthode de collecte des documents, qui s'effectue par soumission d'URL par les webmasters, par des surfeurs-recenseurs, par un crawler,...

- **Méthode d'indexation des documents**

Méthode d'indexation

Indexation manuelle ou automatique

Zone de l'indexation

Tous les éléments de la structure de la page qui sont repris pour l'indexation sont indiqués ici (title, meta, keywords,...)

RECHERCHE

- **Méthode de recherche**

Méthode de recherche

La méthode de recherche au sein de l'outil de recherche est expliqué ici (navigation dans les menus de l'annuaire, introduction de mots-clés dans le masque de saisie de l'annuaire ou du moteur de recherche,...)

Possibilité d'effectuer une recherche avancée

Explication de la procédure afin d'effectuer une recherche avancée pour les moteurs de recherche

- **Essais de recherche pour les moteurs**

	Mot	Mot avec majuscule	Nom propre	Deux mots espacés	Deux mots inversés	Deux mots reliés (AND)
Terme(s)	paris	Paris	Georges Simenon	journal internet	internet journal	+internet +journal
Résultats						

	Un mot exclu (NOT)	Deux mots proches (NEAR)	Deux mots séparés (OR)	Adjacence (ADJ)	Terme dans le champ titre
Terme(s)	+internet -journal	internet NEAR journal	internet OR journal	moteur ADJ recherche	title:internet
Résultats					

	Troncature automatique à droite	Troncature automatique à gauche	Troncature manuelle à droite	Troncature manuelle à gauche	Troncature manuelle au centre	Mot avec caractères accentués	Mot sans caractères accentués
Terme(s)	mot	eur	mot*	*eur	in*que	liège	liege
Résultats							

Des recherches thématiques peuvent également être menées. Ainsi, des termes fixes seront recherchés dans les différents outils afin de vérifier si le domaine, ou un site en particulier, est indexé ou non dans la base de donnée.

- **Possibilités d'interrogation**

Via les analyses précédentes, il est possible de dresser un tableau reprenant les possibilités d'interrogation du moteur de recherche :

	Recherche
Prise en compte de la Casse	Oui ou Non
Opérateur AND	Oui ou Non
Opérateur OR	Oui ou Non
Opérateur NOT	Oui ou Non
Opérateur NEAR	Oui ou Non
Opérateur ADJ	Oui ou Non
Troncature automatique à gauche	Oui ou Non
Troncature automatique à droite	Oui ou Non
Troncature manuelle à gauche	Oui ou Non
Troncature manuelle au centre	Oui ou Non
Troncature manuelle à droite	Oui ou Non
Prise en compte des caractères accentués	Oui ou Non
Possibilité de recherche dans un champ	Oui ou Non
Lemmatisation* lors de la recherche	Oui ou Non
Ordre des mots	Oui ou Non
Rejet des mots vides (<i>le, un, pour,...</i>)	Oui ou Non

* "La lemmatisation est l'opération consistant à lier un mot au singulier avec son pluriel; un verbe conjugué avec son infinitif; un terme français avec son équivalent anglais,... et dès lors permettant à la recherche d'être plus pertinente."

- **Essais de recherche pour les annuaires**

Les recherches pour les annuaires doivent être des recherches thématiques. Ainsi, des termes fixes seront recherchés dans l'arborescence afin de vérifier si le domaine, ou un site en particulier, est indexé ou non dans la base de donnée.

- **Présentation des résultats pour les moteurs de recherche**

Mode de classement des résultats

Classement par ordre alphabétique, de pertinence, par vote,...

Nom du site

Oui, non

URL du site

Oui, non

Lien vers le site

Oui, non

Mots-clés recherchés mis en évidence

Oui, non

Résumé du document

Oui, non

Extrait du document

Oui, non

Mesure de pertinence

Oui, non

Divers et particularités propres

Toutes les particularités sont citées ici (créateur du document, taille, dernière mise à jour, dernière visite, extensions présentes,...)

- **Présentation des résultats pour les annuaires**

Mode de classement des résultats

Classement par ordre alphabétique, de pertinence, par vote,...

Nom du site

Oui, non

URL du site

Oui, non

Lien vers le site

Oui, non

Résumé du document

Oui, non

Extrait du document

Oui, non

Divers et particularités propres

Toutes les particularités sont citées ici (créateur du document, taille, dernière mise à jour, dernière visite, extensions présentes,...)

DESIGN DU SITE ET ARCHITECTURE

- **Design du site**

Présence de publicités

Oui, non

Régie publicitaire

Nom de la régie publicitaire

Lisibilité

Oui, non

Facilité de navigation

Oui, non

Rédaction en français/néerlandais/anglais correct

Niveau de la qualité de rédaction du site en français, en néerlandais ou anglais

- **Architecture html et référencement**

Présence de métadonnées

Oui, non, et lesquelles. Ces métadonnées proviennent de la page principale (cf. supra : l'URL de la page d'accueil)

Le site est-il indexé dans

Google A titre indicatif, le site est-il repris dans les bases de données de Alltheweb l'annuaire (Yahoo! International et Yahoo! France) et des moteurs de Yahoo! recherche (Google et Alltheweb) les plus performants actuellement

Nombre de pages pointant vers le site

*A titre purement indicatif, combien de pages web pointent vers le site, via la balise **link**: dans Google. Ceci permet de connaître la « popularité » du site, afin de savoir combien de sites parlent de l'outil analysé,...*

Exemples de tableaux de synthèse

Une fois la grille complétée, différents tableaux de comparaison des outils analysés peuvent être facilement et rapidement créés, permettant ainsi d'avoir une vue d'ensemble de tous les outils afin de pouvoir être capable de les comparer. Voici des exemples de tableaux de synthèse provenant d'analyses effectuées préalablement pour huit outils de recherche belges et six outils de recherche mondiaux. Les outils de recherche analysés sont les suivants :

Alltheweb

<http://www.alltheweb.com>

Moteur de recherche international, Alltheweb peut être considéré comme le frère d'Altavista. Son but est d'indexer le maximum de type de documents (pages html, mais aussi fichiers audios, vidéos, images,...).

Altavista International

<http://www.altavista.com>

Moteur de recherche international, Altavista est l'un des pionniers des moteurs de recherche. Présent sur l'Internet depuis de nombreuses années, mais en perte de public face à Google, Altavista se concentre actuellement uniquement sur la recherche, laissant de côté ses projets de portails et de messagerie.

Belgian Sites

<http://belgiansites.org>

Belgian Sites est un site non officiel géré par un particulier afin de créer un annuaire des sites officiels belges. La qualité de la base est très bonne, le site accessible en français et néerlandais, mais le design plus que pesant rend le site pénible à consulter...

Charline

<http://www.charline.be>

Portail des villes francophones de Belgique, Charline est un site offrant différents accès aux grandes cités francopho-

nes de Wallonie et Bruxelles. L'idée est intéressante, mais le résultat peu pratique à utiliser, les liens brisés n'étant pas remis à jour, et les annuaires ne possédant que peu de sous catégories.

Dmoz International

<http://dmoz.org>

L'idée première de Dmoz est de créer une multiplicité d'annuaires pour différentes parties de la planète. Cependant, le fonctionnement de Dmoz est différent des autres annuaires : Dmoz est basé sur le bénévolat : chaque internaute peut devenir gestionnaire d'une catégorie, et donc, se consacrer uniquement à l'indexation des sites suggérés par d'autres internautes.

Gojapser

<http://www.gojasper.be>

Outil de recherche belge, Gojapser offre un annuaire très complet, en français et en néerlandais. Il permet de trouver rapidement une information recherchée sur l'Internet.

Google International

<http://www.google.com>

Google est actuellement l'un des moteurs de recherche internationaux les plus puissants. Son atout : permettre des recherches dans des pages qui ne sont plus présentes sur le Net, mais que Google a mémorisé dans sa mémoire, jusqu'à la prochaine indexation de la page par le crawler. Il est donc possible de retrouver une information qui a disparu depuis peu d'un site via Google.

Jaan

<http://www.jaan.be>

Annuaire et moteur de recherche belge néerlandophone, Jaan est un des seuls véritables moteurs de recherche belges actuellement. Portail pour la Flandre et les Pays-Bas, le site est avant tout un portail, mais le moteur de recherche offre généralement de bons résultats pour la Flandre. L'annuaire cependant, bien que contenant de nombreux sites indexés,

ne possède qu'une hiérarchie trop peu détaillée.

Le Potin

<http://www.lepotin.com>

Le Potin franco-belge du Web est principalement un annuaire belge francophone, combiné à un moteur de recherche, qui offre une multitude de services divers. Le site est principalement divisé en 5 catégories principales, toutes basées sur des statistiques, résultats des votes des utilisateurs.

Linkcity

<http://www.linkcity.be>

Annuaire belge géré par des bénévoles, Linkcity est basé sur la même idée que Dmoz : des bénévoles gèrent une ou plusieurs rubriques de l'annuaire afin de posséder le maximum de liens, recherchés par les bénévoles, ou soumis par les internautes. Grand atout de Linkcity : les pages sont entièrement paramétrables par les utilisateurs de l'annuaire.

Lycos International

<http://www.lycos.com>

Autre géant sur l'Internet, Lycos propose des portails spécifiques pour plusieurs zones géographiques. Le portail est en lui-même très complet. Cependant, pour la recherche, Lycos Int. utilise la base

de données de Fast pour le moteur de recherche, et pour l'annuaire, celle de Dmoz.

Search Belgium

<http://www.search-belgium.com>

Portail belge présentant un annuaire combiné à un moteur de recherche, Search Belgium est un outil très agréable à utiliser, et offrant de nombreux résultats pour chaque recherche. De plus, le site propose des interfaces en français, néerlandais et anglais.

Webwatch

<http://www.webwatch.be>

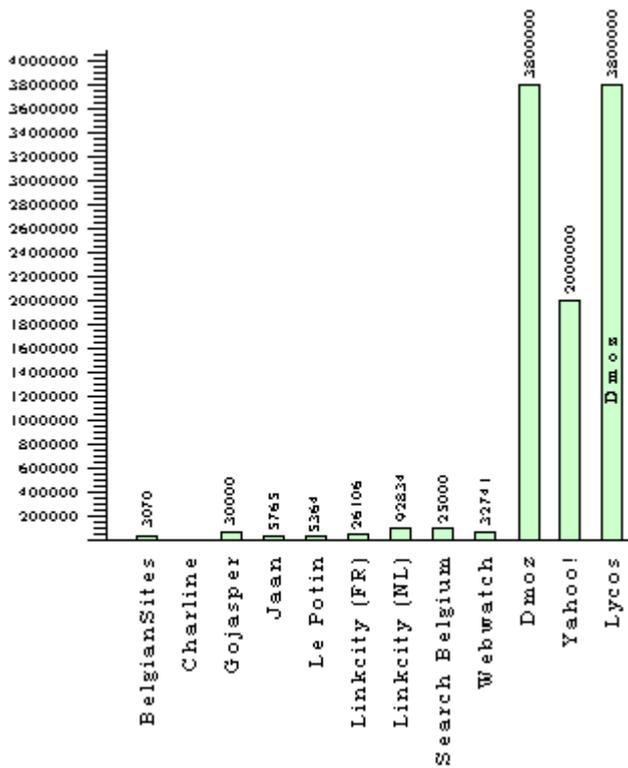
Webwatch est un incontournable sur le Net belge. Portail combinant moteur de recherche et annuaire, Webwatch est un outil indispensable pour les recherches spécifiques à la Belgique, et ce, via une interface en français, néerlandais ou anglais.

Yahoo ! International

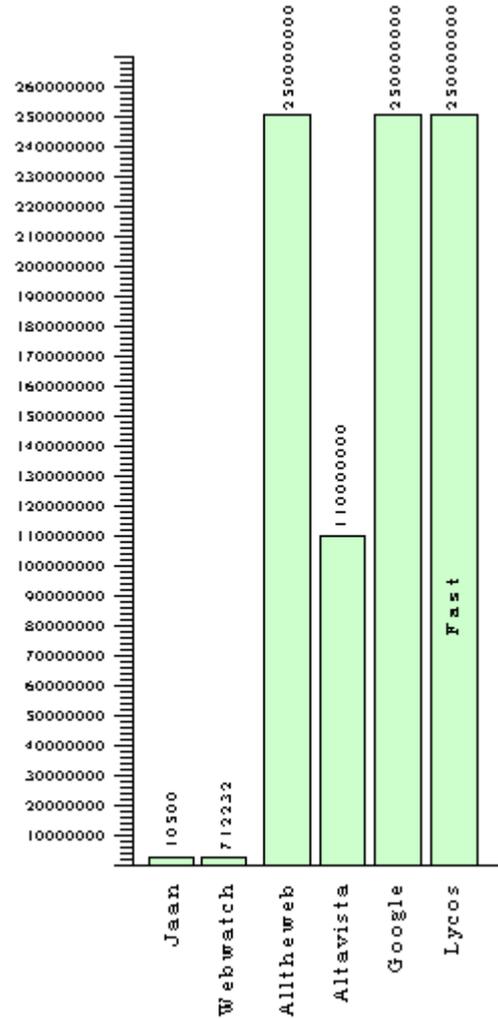
<http://www.yahoo.com>

Pionnier sur l'Internet en matière de recherche dans des annuaires, Yahoo ! est le plus ancien annuaire présent sur le Net, offrant des annuaires spécifiques par zones géographiques dans le monde. Yahoo ! permet également des recherches via un moteur de recherche, celui de Google.

Nombre de sites indexés
pour les annuaires



Nombre de pages indexées
pour les moteurs de
recherche



Source : données fournies par les webmasters au mois de septembre 2002 et le site Abondance : <http://www.abondance.com/>

Remarque : Après plusieurs envois d'e-mails à Charline, aucune réponse ne fut reçue concernant la taille de leur base de données

Généralités des outils de recherche

	Informations générales			Langues du site			Type d'outils		
	Site à nature commerciale	Date création	Aide en ligne	Français	Néerlandais	Anglais	Moteur de recherche	Annuaire	Portail
Belgian Sites		2000	■	■	■		(■)	■	
Charline	■			■			(■)	■	■
Gojasper	■	1999			■		(■)	■	
Jaan		2000	■		■		■		■
Le Potin		1998		■				■	
Linkcity	■	2001	■	■	■		(■)	■	
Search Belgium	■	2000		■	■	■	(■)	■	
Webwatch	■	1996	■	■	■	■	■	■	■
<i>Alltheweb</i>	■	1999	■			■	■		
<i>Altavista International</i>	■	1995	■			■	■	Look smart	■
<i>Dmoz International</i>	■	1998	■			■	(■)	■	
<i>Google International</i>	■	1998	■			■	■	Dmoz Google	
<i>Lycos</i>	■	1994	■			■	■	Fast Dmoz	■
<i>Yahoo International</i>	■	1994	■			■	(■)	■	■

L'icône (■) signifie qu'un moteur de recherche est présent sur le site afin d'effectuer une recherche sur les pages du site, mais n'est donc pas un réel moteur de recherche en soi

Couverture et indexation des sites

	Méthode d'indexation		Méthode de récolte des documents			Couverture				Autres ressources
	Indexation manuelle	Indexation automatique	Soumission des URL	Modification de l'indexation possible	Modification protégée	Web belge	Web belge francophone	Web belge néerlandophone	Web mondial	
Belgian Sites	■		■			■				
Charline	■		■				■			
Gojasper	■		■			■				
Jaan		■	■	□				■		
Le Potin	■		■	■					■	Divers annuaires
Linkcity	■		■	■		■				
Search Belgium	■		■	■		■				
Webwatch	■	■	■	■□	■	■				Usenet belges
<i>Alltheweb</i>		■	■	□				■		Nouvelles, images, vidéos, MP3, fichiers FTP
<i>Altavista International</i>		■	■	□				■		Images, fichiers audios et vidéos, nouvelles
<i>Dmoz International</i>	■		■	■				■		
<i>Google International</i>		■	■	□				■		Images, usenet, nouvelles
<i>Lycos International</i>	!	!	!	!	!			■		
<i>Yahoo International</i>	■		■	■				■		

L'icône (☐) signifie qu'une modification est automatique pour les moteurs de recherche via une simple nouvelle soumission de l'URL

Concernant *Lycos*, l'annuaire et le moteur de recherche étant gérés par d'autres sociétés que *Lycos*, les soumissions et modifications doivent s'effectuer via *Dmoz* ou *Fast (Alltheweb)*.

Possibilités de recherche pour les moteurs de recherche

	Prise en compte de la casse	Opérateur AND	Opérateur OR	Opérateur NOT	Opérateur NEAR	Opérateur ADJ	Troncature automatique à gauche	Troncature automatique à droite	Troncature manuelle à gauche	Troncature manuelle au centre	Troncature manuelle à droite	Prise en compte des caractères accentués	Possibilité de recherche dans un champ	Lemmatisation lors de la recherche	Correction orthographique	Rejets des mots vides	Ordre des mots
Jaan	■	■		■			■	■			■	■					■
Webwatch	■	■	■	■	■	■	■	■	■	■	■	■					■
<i>Alltheweb</i>		■		■									■				
<i>Altavista Int.</i>		■	■	■	■						■	■			■	■	
<i>Google Int.</i>	■	■		■								■	■		■	■	
<i>Lycos Int.</i>	<i>Voir Fast (Alltheweb)</i>																

CONCLUSIONS

L'Internet est un outil de travail plus qu'intéressant, si l'on est capable de bien l'utiliser. Il devient un outil de recherche de plus en plus courant dans la vie de tous les jours de plusieurs millions de personnes, tant au niveau privé qu'au niveau professionnel. Mais l'Internet possède ses bons et mauvais cotés. Ses ressources ne cessent de croître de façon exponentielle, et le traitement de cette masse d'information est difficile à réaliser. Ainsi, l'on trouve sur le Net un peu de tout en matière d'outil de recherche ; des moteurs de recherche mondiaux faisant partie de multinationales, aux annuaires mondiaux réalisés par des bénévoles, en passant par de petits annuaires thématiques gérés par un webmaster passionné. L'intérêt de ces outils pour l'utilisateur varie fortement de l'un à l'autre. Aussi faut-il dès lors procéder à une analyse en profondeur de l'histoire de l'outil, afin d'en connaître ses capacités exactes. Dès lors, l'utilisation de ces grilles de façon systématique permettra de retirer les principaux fondements des outils de recherche, ainsi que leur

évolution sur ce qui est le marché de la recherche sur le Net. Ces grilles permettent en permanence de comparer à tout moment les capacités de recherche et l'évolution d'un outil donné, et sa compétitivité avec des outils semblables.

Dans les exemples d'analyses donnés ci-dessus, on peut remarquer qu'Altavista est à la traîne en matière de nombre de documents indexés dans sa base de données par rapport à Google ou à la base de données Fast, utilisée par Alltheweb et Lycos. Point de vue annuaires, Yahoo! se fait également dépasser par Dmoz. De même, en matière de possibilité de recherche sur les moteurs de recherches belges, on peut facilement remarquer que la base de Webwatch est plus fournie que celle de Jaan, et de même, offre plus d'opérateur de recherche. Au niveau international par contre, Altavista se rattrape sur Google en offrant un peu plus d'opérateurs.

Ces grilles permettront aux utilisateurs de mettre en avant ce qu'ils souhaitent en matière de qualité de la recherche sur le

Web, tous les critères pouvant être pris en compte, ou seulement certains, en fonction du public qui utilisera ces outils, le but premier de ces tableaux restant de guider les internautes vers des outils de recher-

ches correspondant à leurs besoins, souhaits et attentes.

Annexe : exemple d'analyse d'un annuaire : Linkcity



Linkcity

Annuaire et moteur de recherche internet

26 106 sites pour la version française

42 834 sites pour la version néerlandophone



Annuaire belge géré par des bénévoles, Linkcity est un annuaire qui ne cesse de se développer, offrant des recherches de qualité, vu la richesse de sa base, et l'inexistence de liens cassés. Grand atout de Linkcity : ses pages sont entièrement paramétrables par les utilisateurs de l'annuaire. LinkCity est accessible via sa propre URL propre mais aussi intégré en tant qu'index officiel dans des portails tels que Skynet, SwinG, Infobel, Planet Internet, Rendez-Vous, NetBel, CyberCafé21, etc.

DESCRIPTION

- **Généralités**

Nom

Linkcity : bibliothèque de liens

Type

Annuaire, avec moteur de recherche interne à l'annuaire

Nature du site

Site à nature commerciale, mais géré par des bénévoles

URL de la page principale

<http://www.linkcity.be/>

URL de la page à partir de laquelle une recherche peut être effectuée

En français : <http://www.linkcity.be/scripts/FR/Page.cfm?xcookie=N>

En néerlandais : <http://www.linkcity.be/scripts/NL/Page.cfm?sourcecheck=>

URL de soumission de site

Les sites doivent être proposés aux gestionnaires de pages via un formulaire qui met en contact l'internaute et celui-ci. Les sites doivent être soumis à partir du lien présent sur la page de la rubrique qui accueillera le site

Couverture

Web belge en majorité, mais également des liens internationaux. De plus, une version de Linkcity pour la France existe (<http://fr.linkcity.com/>)

Langue(s) du site

Français, néerlandais

Date de création

31 janvier 2001

Dernière modification

Pas de date indiquée

- **Particularités**

Auteurs et webmasters du site

Arnaud Huret et Patrick de Schutter pour le concept ; Gilles Renard pour le contenu ; et Daniel Renotte pour la programmation

Nombre d'employés

Un employé, plus les gestionnaires bénévoles des catégories

Données de contact du site

*Linkcity – NetCapital info@linkcity.be
55 rue des anciens étangs
1190 Bruxelles*

Nombre de sites indexés

26 106 pour la version française ; 42 834 pour la version néerlandophone

Fréquentation du site

Statistiques non consultables sur le site

Périodicité de mise à jour

Sites encodés très rapidement dans la base de données, vu le nombre de bénévoles travaillant pour Linkcity

RECOLTE ET INDEXATION DES DOCUMENTS

- **Documents indexés**

Type de documents indexés

Le Web

Langue(s) des documents indexés

Français, anglais, néerlandais, allemand

- **Méthode de récolte des documents**

Collecte des documents

Les webmasters soumettent des URL aux gestionnaires de pages, qui peuvent également enrichir leur(s) page(s) en intégrant eux-mêmes des sites qu'ils sélectionnent

- **Méthode d'indexation des documents**

Méthode d'indexation

Indexation manuelle

Zone de l'indexation

Les zones indexées sont le nom des pages thématiques, le nom des catégories, et le nom des liens

Remarque : la base de données de Linkcity est capable de signaler les liens qui ne sont plus actifs afin de les enlever de la base de données, et ainsi conserver une banque de liens à jour

RECHERCHE

- **Méthode de recherche**

Méthode de recherche

La recherche s'effectue via navigation dans l'annuaire, ou via un moteur qui recherche dans les sites de l'annuaire

Possibilité d'effectuer une recherche avancée

Non

- **Essais de recherche pour les moteurs**

	Mot	Mot avec majuscule	Nom propre	Deux mots espacés	Deux mots inversés	Deux mots reliés (AND)
Terme(s)	paris	Paris	Georges Simenon	Wallonie presse	presse Wallonie	+Wallonie +presse
Résultats	301	301	1	1	0	0

	Un mot exclu (NOT)	Deux mots proches (NEAR)	Deux mots séparés (OR)	Adjacence (ADJ)	Terme dans le champ titre
Terme(s)	+Wallonie -presse	Wallonie NEAR presse	Wallonie OR presse	moteur ADJ recherche	title:internet
Résultats	0	0	0	0	0

	Troncature automatique à droite	Troncature automatique à gauche	Troncature manuelle à droite	Troncature manuelle à gauche	Troncature manuelle au centre	Mot avec caractères accentués	Mot sans caractères accentués
Terme(s)	mot	eur	mot*	*eur	in*que	liège	liege
Résultats	201	201	201	201	(201)	201	5

- **Possibilités d'interrogation du moteur de recherche interne de Linkcity**

	Recherche
Prise en compte de la Casse	Non
Opérateur AND	Oui
Opérateur OR	Non
Opérateur NOT	Non
Opérateur NEAR	Non
Opérateur ADJ	Non
Troncature automatique à gauche	Oui
Troncature automatique à droite	Oui
Troncature manuelle à gauche	Oui
Troncature manuelle au centre	Non
Troncature manuelle à droite	Oui
Prise en compte des caractères accentués	Oui
Possibilité de recherche dans un champ	Non
Lemmatisation lors de la recherche	Non
Ordre des mots	Oui
Rejet des mots vides (<i>le, un, pour,...</i>)	Non

- **Essais de recherche pour les annuaires**

Terme(s)	Bibliothèque	Wallonie	Transport
Résultats	Oui	Oui	Oui

- **Présentation des résultats pour les moteurs de recherche**

Mode de classement des résultats

Classement par ordre alphabétique

Nom du site

Oui

URL du site

Non

Lien vers le site

Oui

Mots-clés recherchés mis en évidence

Non

Résumé du document

Oui

Extrait du document

Non

Mesure de pertinence

Non

Divers et particularités propres

/

- **Présentation des résultats pour les annuaires**

Mode de classement des résultats

Classement par ordre alphabétique

Nom du site

Oui

URL du site

Non

Lien vers le site

Oui

Résumé du document

Oui

Extrait du document

Non

Divers et particularités propres

/

DESIGN DU SITE ET ARCHITECTURE

- **Design du site**

Présence de publicités

Oui

Régie publicitaire

Beweb - <http://www.beweb.com/>

Lisibilité

Oui

Facilité de navigation

Oui

Rédaction en français/néerlandais/anglais correct

Bon niveau de rédaction

- **Architecture html et référencement**

Présence de métadonnées

description, keywords, author, robots, revisit-after, reply-to, language

Le site est-il indexé dans

Google Oui

Alltheweb Oui

Yahoo! Non

Nombre de pages pointant vers le site

58 pages via la balise link ::www.linkcity.be dans Google

* * *