

Les outils de recherche sur l'Internet

Fiche rédigée par Michel Narcy, formateur informatique et rédacteur en chef de Médialog

De nombreux services de recherche sont disponibles sur la Toile, comme Alta Vista ou Yahoo !. Ces services utilisent des moteurs de recherche.

Un moteur d'indexation et de recherche est un outil qui permet d'extraire d'une information, principalement textuelle, les mots ou termes qui la représentent au mieux et de les stocker dans un index ; le même outil parcourt ensuite cet index afin d'identifier les termes les plus pertinents par rapport à ceux de la question de l'utilisateur, puis de trier les informations à lui fournir en retour.

Les sites de recherche sur la Toile (généralement appelés moteurs de recherche par confusion entre le service et l'outil utilisé) se divisent en deux catégories : les annuaires ou répertoires (indexation humaine) et les robots (indexation automatique).

1. Les annuaires

Les annuaires ou répertoires thématiques (aussi appelés catalogues ou arborescence) présentent un ensemble de domaines familiers sur lesquels il suffit de cliquer pour afficher plusieurs sous-domaines. En cliquant sur un sous-domaine, on accède à un ensemble de sous-domaines afférents, et ainsi de suite jusqu'à atteindre la page Web recherchée.

Les annuaires sont utilisés pour trouver un site spécifique parlant d'un thème donné. Ils ressemblent aux « pages jaunes » qui se seraient structurées à l'aide d'un thésaurus interactif.

L'indexation d'un annuaire est, en grande partie, humaine et donc assortie de mots-clés (catégories thématiques) et de résumés.

Les annuaires de recherche ont des avantages et des inconvénients.

Avantages :

- La facilité d'utilisation : il suffit de développer une arborescence pour trouver les pages Web en rapport avec le sujet sélectionné.
- Le faible nombre de pages sélectionnées pour chaque catégorie.
- La qualité du contenu : le nombre de sujets traités dans un répertoire de recherche étant limités, et les informations organisées par des opérateurs humains, les documents référencés sont souvent de bonne qualité et sont effectivement en rapport avec le sujet sélectionné.

Inconvénients :

- L'organisation thématique des pages Web peut rendre ardue la localisation de sites dont le sujet est difficile à classer.
- L'agencement des annuaires reposant sur des opérateurs humains, leur mise à jour est fatalement plus lente que celle des index de recherche, entièrement alimentés par des robots.

Le plus célèbre des annuaires est Yahoo ! D'autres annuaires francophones :

En version anglaise : www.yahoo.com
En version française : www.yahoo.fr

Lycos www.lycos.fr
Nomade : www.nomade.fr
Ecila : www.ecila.fr
Lokace : www.lokace.fr
Francité : www.francite.com

2. Les moteurs

Les moteurs de recherche fonctionnent différemment des annuaires de recherche. En entrant un ou plusieurs mots dans la zone de texte du moteur, on obtient en retour une liste plus ou moins importante de sites en rapport avec le ou les sujets spécifiés.

L'indexation d'un moteur est automatique.

Les moteurs d'indexation et de recherche peuvent être répartis en deux groupes : *full-text* et *keyword*.

Les moteurs *full-text* mémorisent la totalité des mots contenus dans les pages Web. Les moteurs *keyword* se contentent de mémoriser les mots et groupes de mots « importants ». Les moteurs *full-text* renvoient un nombre important de réponses, les moteurs *keyword* renvoient un nombre de réponses plus « raisonnable ». En contrepartie, tous les documents qui traitent un sujet donné ne sont pas référencés.

Par exemple, AltaVista et Open Text sont des moteurs full-text. Lycos et Excite sont des moteurs keyword.

Tout comme les annuaires, les moteurs ont leurs avantages et leurs inconvénients.

Avantages :

- Le nombre de pages accessibles est bien plus important que dans un annuaire de recherche.
- Les requêtes envoyées à un moteur de recherche peuvent être affinées en utilisant un véritable langage spécifique à chaque moteur.
- Les moteurs de recherche mettent à jour régulièrement leur base de données. Les sites qui y sont référencés sont donc en général toujours d'actualité.

Inconvénients :

- Il est nécessaire de se familiariser avec le langage de requêtes propre à chaque moteur.
- Le nombre de sites sélectionnés est bien souvent trop important.
- Étant donné que les moteurs sont entièrement alimentés par des robots, aucun jugement de valeur n'est appliqué sur le contenu de chaque site.

Les moteurs de recherche les plus utilisés sont :

| En anglais : | | En français | |
|-------------------|--|-------------------|--|
| Alta Vista | www.altavista.com | Alta Vista | www.altavista.fr |
| Infoseek | www.infoseek.com | Lycos | www.lycos.fr |
| Hotbot | www.hotbot.com | Excite | www.excite.fr |
| Google | www.google.com | Voila | www.voila.fr |