

## Annuaire ou moteurs de recherche ? Quelles différences ?

### Annuaire (guides, répertoires)

Un annuaire propose une sélection généraliste de **sites web** qu'une équipe de collaborateurs décrit au travers de fiches comportant, en règle générale, un titre, l'adresse du site (URL) et un bref **commentaire** d'une longueur dépassant rarement 15 à 25 mots.

Les annuaires peuvent être considérés comme des répertoires ("pages jaunes" du Web) bâtis par des êtres humains ("netsurfeurs", "documentalistes"... ) qui :

- choisissent les sites à répertorier en parcourant le Web (ou vérifient les sites qui demandent/paie leur inscription)
- rédigent la fiche descriptive correspondante
- inscrivent chaque site au sein de l'annuaire dans une ou plusieurs **catégorie(s)**.

Les annuaires peuvent être **parcourus par leur arborescence de classement**, de rubriques générales en sous-rubriques plus précises. Les annuaires disposent généralement d'un système de recherche **interne** où l'internaute saisit un mot-clé dans un formulaire. La **recherche des occurrences de ce terme s'effectue dans les fiches descriptives des sites inscrits ou dans les noms des rubriques de l'annuaire, voire dans un thésaurus mais PAS dans le contenu réel des pages des sites.**

Exemple d'annuaire : dmoz <http://www.dmoz.fr>

### Moteurs de recherche

Un moteur de recherche fonctionne sur un système radicalement différent de celui de l'annuaire. Des **robots logiciels** (appelés *crawlers*, *spiders*...) parcourent le Web, vont de page en page (de lien en lien), **sauvegardent et indexent** au fur et à mesure de leurs « visites » une partie du contenu des pages rencontrées, constituant ainsi un "index". Cette collection de **copies de pages Web est interrogeable via des requêtes**. Les « index » des moteurs contiennent les copies de milliards de pages ! Le robot logiciel repasse « sur le Web » selon un délai plus ou moins long pour **actualiser automatiquement sa base de données** : découvrir de nouvelles pages, supprimer celles qui n'existent plus, mettre à jour celles qu'il a indexées au préalable et qui ont changé (en sauvegarder une version plus récente). On dit alors qu'il « rafraîchit sa base de données » (son index).

Les moteurs de recherche disposent d'un système de recherche où l'internaute saisit un mot-clé dans un formulaire. Le moteur va en rechercher les occurrences du mot-clé dans son index, c'est-à-dire **dans le « contenu réel » des copies de pages Web sauvegardées**. Après détermination du "lot" de pages contenant le mot-clé, le moteur affiche les pages trouvées dans un certain ordre, selon un algorithme de pertinence complexe qui n'est pas explicitement divulgué.

Exemple de moteur : Google <http://www.google.fr>

### Outils de recherche mixtes

Les outils de recherche s'enrichissent de **services annexes**. Proposer des sources d'informations comme une encyclopédie, des dépêches d'actualité, un espace météo ou "shopping"... attire l'internaute vers un outil de recherche (marketing).

Pour compliquer le tout (!) la plupart des outils de recherche sur le Web (Google, Yahoo, Altavista, etc.) **proposent à la fois un annuaire et un moteur de recherche** afin notamment d'éviter que les recherches adressées aux annuaires restent sans réponse. **Identifier si l'on interroge un annuaire ou un moteur de recherche est important** mais pas toujours évident à déceler ! Un annuaire est parfois nommé « répertoire », « guide », « sites classés ». Il est nécessaire d'afficher certaines options ou la recherche avancée pour savoir « à quel type d'outil » on a affaire !

		Comparatif	
		Annuaire	Moteurs de recherche
<b>Avantages</b>	<ul style="list-style-type: none"> <li>- Sites choisis et sélectionnés par une équipe humaine</li> <li>- Nombre de sites référencés restreint donc moindre risque de « se perdre »</li> <li>- Sujets généraux faciles à retrouver dans l'arborescence</li> <li>- Cohabitation de deux modes de recherche : par parcours de l'arborescence ou par mot-clé</li> <li>- Association automatisée par l'annuaire de certains mots-clefs des requêtes à d'autres mots au sein d'un thésaurus (qui n'est pas divulgué)</li> </ul>	<ul style="list-style-type: none"> <li>- Grande masse d'informations (abondance et diversité)</li> <li>- Accès à des informations sur des sujets précis</li> <li>- Classement des réponses par « pertinence »</li> <li>- Accès à des pages récemment publiées ou mises à jour (base réactualisée automatiquement)</li> <li>- Utilisation possible d'équations de recherche complexes</li> <li>- Grande efficacité de recherche lorsque l'on maîtrise l'outil</li> <li>- Pas de filtrage humain donc accès à toutes les informations</li> </ul>	
<b>Inconvénients</b>	<ul style="list-style-type: none"> <li>- Nombre de sites référencés restreint donc risque de ne pas trouver l'information recherchée</li> <li>- Ne convient pas aux sujets trop pointus</li> <li>- Risque de ne pas interroger certains sites récents</li> <li>- Recherche par mot-clef parfois moins efficace du fait que le contenu même de la page n'est pas interrogé</li> <li>- Présence de sites qui ont acheté leur place dans l'annuaire</li> </ul>	<ul style="list-style-type: none"> <li>- Grande masse d'informations d'où des URL brisées... et trop d'informations !</li> <li>- Syntaxe des équations de recherche différente d'un moteur à l'autre</li> <li>- Relative opacité des critères de l'ordre d'affichage des sites trouvés (affichage pourtant souvent pertinent...)</li> <li>- Pas de filtrage humain donc accès à des informations de toute nature (être d'une grande vigilance)</li> <li>- Nécessité de rédiger finement les termes de la requête</li> <li>- Présence de sites qui ont acheté leur place dans les pages de résultats de l'annuaire</li> </ul>	