

La recherche sur Internet

Internet se présente un peu comme une bibliothèque publique où chacun peut

- venir **consulter des ouvrages** et
- **déposer des travaux** qu'il a réalisés.

Le problème, c'est qu'il n'y a **pas de** bibliothécaire ! Il n'existe donc pas de catalogue officiel. Personne ne sait donc quels sont les documents qui peuvent être consultés.

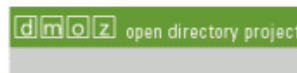
La réalisation d'un catalogue complet des ressources disponibles sur l'Internet est un travail impossible. Certaines sociétés se sont cependant lancées dans l'entreprise de la réalisation de catalogues aussi étendus que possible.

Selon la méthode de travail utilisée par ces sociétés, on parlera plutôt d'**annuaire** ou de **moteur de recherche**.

Un annuaire fournit une liste des sites, classés par thèmes, un peu comme l'annuaire du téléphone.

Les sites référencés sont tous proposés par leur concepteur appelé "webmaster. Un moteur de recherche est une gigantesque banque de données mise à jour continuellement par des ordinateurs "robots" qui parcourent le web. Ces robots parcourent les pages pour constituer une liste des mots figurant dans chaque site.

Quelques exemples :



Fonctionnement des moteurs de recherche

Lorsqu'on effectue une recherche par mots-clés sur un moteur de recherche, celui-ci ne parcourt pas l'intégralité du Web au moment de l'interrogation : cela lui prendrait beaucoup trop de temps.

Phase 1.

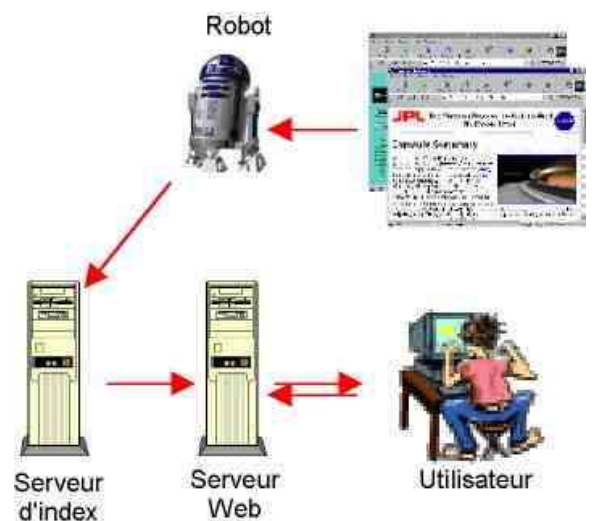
Les moteurs de recherche sont associés à des ordinateurs "robots" d'exploration qui parcourent le Web en permanence en suivant les liens hypertexte.

Phase 2

Les pages "lues" par le robot sont envoyées vers un autre ordinateur qui élabore un **INDEX** à partir des pages lues (comme un livre de référence est souvent muni d'un index). L'intégralité des pages lues n'est donc généralement pas conservée ; seul l'index existe en permanence.

Phase 3

Lorsqu'un utilisateur envoie une requête au **serveur Web** du moteur de recherche, celui-ci interroge le serveur d'index qui lui renvoie les informations demandées.



Fonctionnement des annuaires

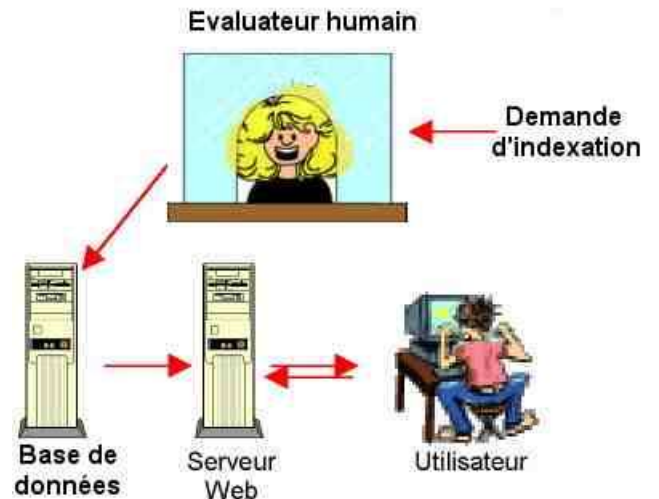
Lorsqu'on effectue une recherche par mots-clefs sur un annuaire, il ne parcourt pas non plus l'intégralité du Web au moment de l'interrogation : cela lui prendrait beaucoup trop de temps.

Phase 1.

Les annuaires reçoivent des demandes de référencement de la part des responsables des sites web qui le souhaitent.

Phase 2

Les demandes sont traitées par un évaluateur humain qui parcourt les sites web proposés. Cet évaluateur juge de l'intérêt du site, vérifie que la demande de référencement est correcte (dans la bonne catégorie) puis accepte ou refuse le référencement. Le référencement peut éventuellement être payant. Le site est alors indexé dans une base de données, un gigantesque catalogue de sites qui pourra être consulté par le web.



Phase 3

Lorsqu'un utilisateur envoie une requête au serveur Web de l'annuaire, celui-ci interroge sa base de données qui lui renvoie les informations demandées.

Exemple de recherche par catégories

La recherche d'un article dans un supermarché peut se faire en parcourant tous les rayons jusqu'au moment où l'article est trouvé. Imaginons que nous cherchons des **olives vertes en conserve**.

Nous savons que nous devons nous diriger vers la partie "**alimentation**" du supermarché.

Nous nous dirigeons alors vers les rayons consacrés à l'"**épicerie**".

Dans ces rayons, ce sont les "**conserves**" qui doivent retenir notre attention.

Le supermarché vend plusieurs sortes de conserves. Dirigeons-nous vers les olives.

Là, nous pouvons nous servir.

Le chemin qui nous a permis de trouver les olives pourrait s'écrire de la manière suivante :

Alimentation > Épicerie > Conserves > Olives

Fiabilité des informations sur le Web

Remarque : il n'existe sur le Web aucune garantie de véracité des informations qui sont déposées.

Imaginons que vous ayez créé un site sur les dinosaures à sang chaud. Vous référencez votre site sur un ou plusieurs moteurs de recherche et annuaires afin de le faire connaître.

Il est peu probable qu'il existe parmi les employés de "Yahoo!" ou de "Voilà" ou d'autres, un spécialiste des dinosaures. Personne ne pourra donc juger de la crédibilité des informations que vous déposez.

Les robots n'ayant pas accès au sens des mots qu'ils détectent ne peuvent pas détecter le sujet d'un document par le simple examen des mots qu'il contient.

Un document contenant le mot "Diplodocus" n'est pas forcément en rapport avec les dinosaures.

Trouver une information intéressante sur Internet n'est pas toujours évident.

Il faut que :

- la personne qui est à la source de l'information ait fait enregistrer son travail dans l'annuaire ;
- la personne qui cherche l'information effectue correctement sa recherche ;
- les informations déposées soient correctes; il faut donc se donner des moyens pour vérifier la crédibilité de la source d'information