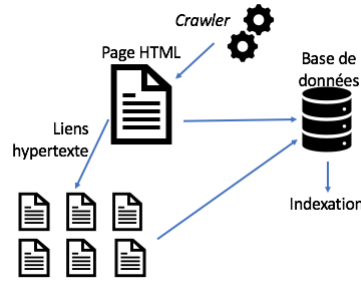


Les moteurs de recherche

Le fonctionnement d'un moteur de recherche se décompose en trois processus principaux :

- ▶ L'exploration
- ▶ L'indexation
- ▶ L'algorithme de recherche



- ▶ **L'exploration** : le web est systématiquement exploré par un robot (appelé crawler ou spider) suivant tous les hyperliens qu'il trouve pour analyser les éléments significatifs des pages WEB : Le code source HTML, l'URL, les images et le nom qui leur est donné, les liens internes et externes, les pages sur lesquelles pointent ces liens, etc.
- ▶ **L'indexation** : ce robot indexe alors les ressources récupérées, en ciblant les mots considérés comme significatifs : Les mots significatifs sont enregistrés dans une base de données et ils leurs sont associés à un poids. Le robot ne tient pas compte non significatifs tels que " de ", " le ", " la ", etc.
- ▶ **La recherche dans le moteur** : un algorithme de recherche est appliqué pour identifier dans le corpus documentaire (en utilisant l'index), les documents qui correspondent le mieux aux mots contenus dans la requête, afin de présenter les résultats des recherches par ordre de pertinence supposée.
Le premier site qui apparaît en premier dans le résultat de la recherche est le site qui a la meilleure indexation.
- ▶ **Le PageRank** est un des algorithmes d'analyse des liens de chaque page WEB pour en mesurer la popularité : Plus il y a de liens qui pointent sur une page, plus le PageRank est élevé. il regarde aussi le trafic de la page, le comportement des visiteurs, le nom de domaine.
- ▶ **Le SOE** : Le score de popularité d'un site internet est appelé SEO (Search Engine Optimisation). Il est à opposer au référencement payant proposé qui permet à des liens d'apparaître plus haut dans le moteur de recherche moyennant une somme d'argent.

- ✚ Les 3 premiers résultats d'un moteur de recherche remportent plus de 80 % des clics
- ✚ Le moteur de recherche Google indexe actuellement près 100 milliards de pages (soit 4% !).
- ✚ Il existe des pays dans lesquels Google n'est pas le moteur le plus utilisé : Chine et Russie !