Comment fonctionne un moteur de recherche ?

Retour au sommaire de la lettre

De nombreux articles traitent des dernières innovations dans la recherche sur Internet mais peu abordent, en revanche, le fonctionnement technique des moteurs de recherche. Nous vous proposons ici une analyse globale du fonctionnement des moteurs et des processus qui sont mis en œuvre pour traiter les documents, stocker les informations les concernant et restituer des résultats aux requêtes des utilisateurs.

Un moteur de recherche est un ensemble de logiciels parcourant le Web puis indexant automatiquement les pages visitées. Trois étapes sont indispensables à son fonctionnement :
- la **collecte d'information.**
- l'**indexation** des données collectées et la constitution d'une base de données.
- le **traitement des requêtes**, avec en particulier un système d'interrogation de la base de données et de classement des résultats en fonction de critères de pertinence.

Deux principaux types de contenus sont actuellement affichés par les moteurs dans leurs pages de résultats :
- les **liens** "**organiques**"ou "naturels", obtenus grâce au *crawling* du Web.
- les **liens sponsorisés.**

Nous allons nous concentrer ici en priorité sur les techniques utilisées par les moteurs pour indexer et retrouver des liens "naturels" et nous n'aborderons pas le traitement spécifique des liens sponsorisés.

Nota : Cet article constitue une première approche, la plus globale possible, du fonctionnement d'un moteur de recherche. Bien sûr, ce fonctionnement est souvent bien plus complexe lorsqu'on l'analyse en détail. Certains d'entre vous, déjà familiers du sujet, ou plus orientés vers les aspects techniques, pourront peut-être le trouver "simpliste"... Mais nous envisageons de vous proposer des articles plus précis sur certains points. N'hésitez pas à nous faire savoir si cela vous intéresse et quels "rouages" des moteurs de recherche vous semblent les plus abscons aujourd'hui, nécessitant des informations complémentaires dans les mois qui viennent... Merci !

Technologies utilisées par les principaux moteurs de recherche

En dehors des trois leaders du marché (Google, Yahoo et MSN), de nombreux moteurs n'utilisent pas leurs propres technologies de recherche mais ils sous-traitent cette partie auprès de grands moteurs.

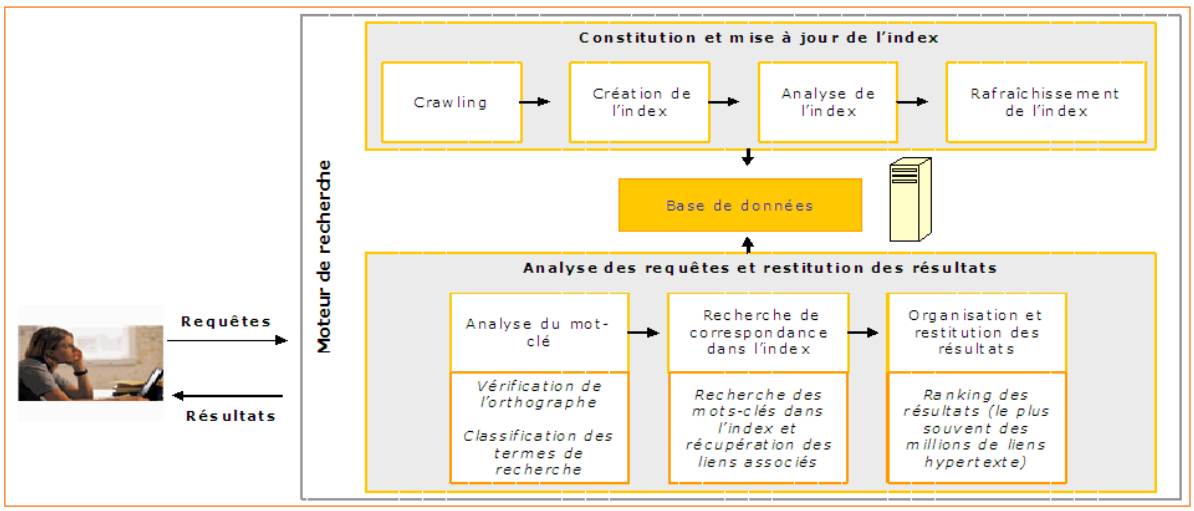
Technologies de recherche actuellement utilisées par les principaux moteurs

		Technologies de recherche							
		Google	YAHOO!	msn	antidot	exalead	mirago	TEOMA	voila
Moteurs anglophones (parts de recherche – Monde - Nielsen Netratings -01/2005)	Google (47%)	X							
	Yahoo (21%)		X						
	MSN (13%)			X					
	AllTheWeb *(Yahoo)*		X						
	A9 *(Amazon)*	X						X	
	AltaVista *(Yahoo)*		X						
	Ask Jeeves							X	
	Eurekster		X						
	Exalead					X			
	Hotbot	X						X	
	Lycos							X	
	Mirago						X		
Moteurs francophones (parts de trafic – France - Weborama - Takezo/Brioude - 01/2005)	Google (80,14%)	X							
	Yahoo (5,59%)		X						
	MSN (3,96%)			X					
	Wanadoo (3,48%)								X
	AOL.FR (2,60%)					X			
	Free (1,96%)	X							
	Tiscali (<0,5%)		X						
	Club Internet (<0,5%)	X							
	Lycos (<0,5%)		X						
	Meceoo *(Abondance)*		X						
	La Poste				X				
	Ujiko *(Kartoo)*		X						

Sources : Abondance - Nielsen Netratings (01/2005) - Weborama - Takezo/Brioude (01/2005)

Principe de fonctionnement d'un moteur de recherche

Pour leur fonctionnement, les moteurs de recherche suivent plusieurs étapes : des robots explorent le Web de lien en lien et récupèrent des informations. Ces informations sont ensuite indexées par des moteurs d'indexation, les termes répertoriés enrichissant un index régulièrement mis à jour (une base de données des termes descriptifs retenus). Une interface de recherche permet enfin de restituer des résultats aux utilisateurs en les priorisant en fonction de leur pertinence. Cette opération est appelée le "ranking".

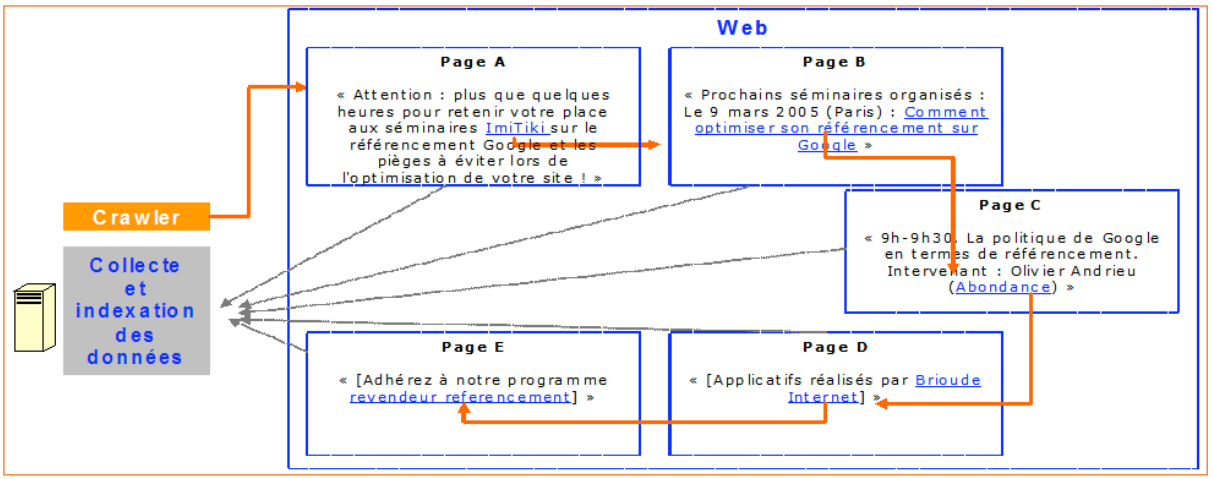


Composants technologiques d'un moteur

Le crawler

Les crawlers (également appelés spiders, robots ou bots) sont des **programmes de navigation visitant les pages Web et leurs liens de manière continue en vue d'indexer leurs contenus**. Ils parcourent les liens hypertextes entre les pages et reviennent périodiquement visiter les pages retenues pour prendre en compte les éventuelles modifications.
Parmi les spiders connus, citons notamment le spider "Googlebot" de Google, "Yahoo ! Slurp" de Yahoo, "Henri Le Robot Mirago" du moteur Mirago ou encore le plus récent "MSNBot" de MSN.

Processus de crawling



Les webmasters ont la possibilité d'indiquer aux spiders dans un fichier "robots.txt" les pages qu'ils souhaitent voir indexer et les pages qu'ils n'entendent pas faire référencer (voir à ce sujet la page http://docs.abondance.com/robots.html).

La plupart des moteurs gère le crawling de manière "différenciée" et non "linéaire". Ils visitent plus fréquemment les sites à fort trafic et à fort taux de renouvellement des contenus et se rendent moins souvent sur les pages "statiques" moins fréquentés. Ainsi, une page qui est mise à jour quotidiennement (par exemple, un site d'actualité) sera visitée chaque jour ou tous les deux jours par le robot tandis qu'une page rarement modifiée sera "crawlée" toutes les quatre semaines en moyenne.

A noter que la technique de suivi des liens hypertextes des crawlers pose plusieurs difficultés pour :
- L'indexation des pages qui ne sont liées à aucune autre (pages non liées) et ne peuvent donc pas être répérées.
- L'indexation des pages "dynamiques" de périodiques ou de bases de données (ces pages étant moins facilement prises en compte).

Le passage des spiders sur les sites peut être vérifié par les webmasters en analysant les fichiers "logs" des sites sur les serveurs (ces fichiers indiquent l'historique des connexions qui ont eu lieu dont celles des spiders). La plupart des outils statistiques comprennent une partie "visites des robots". Attention cependant : ces outils doivent le plus souvent être configurés pour prendre en compte tous les robots, notamment émanant de moteurs français. Les outils statistiques, notamment d'origine américaine, ne prennent pas toujours en compte ces spiders "régionaux"...

Plusieurs applications en ligne permettent également d'analyser les visites des robots sur des pages données (voir notamment les solutions gratuites http://www.robotstats.com/ et http://www.spywords.com/). Des marqueurs doivent être intégrés par les webmasters dans les pages et les services surveillent si l'un des visiteurs est le robot d'un moteur de recherche.

Le moteur d'indexation

Le spider envoie au moteur d'indexation les informations collectées.

Plusieurs systèmes d'indexation des données sont alors utilisés :

- **Indexation des mots-clés**, en exploitant les balises meta (meta-tags) insérées par les webmasters dans le code source des pages html, balises qui comprennent entre autres le résumé et les mots-clés attribués par l'auteur à la page.
- **Indexation des titres** (informations qui sont de moins en moins utilisées car les titres des documents ne reflètent pas toujours le contenu de la page) **ou de quelques lignes des documents.**
- **Indexation en texte intégral** (c'est le cas le plus fréquent, tous les mots d'une page sont alors indexés).

Le plus souvent, les systèmes d'indexation se chargent d'identifier en "plein texte" l'ensemble des mots des textes gérés par le moteur ainsi que leur position.

Seule une partie des pages collectées sont conservées à terme par les moteurs. Les moteurs utilisent en effet plusieurs critères tels que la richesse du contenu texte et la lisibilité du contenu par le spider, l'adéquation entre les mots clés présents dans les balises meta et le contenu des pages pour décider des sites qui seront indexés. De même, des pages à contenu trop proches sont souvent "dédoublonnées" (phénomène de "duplicate content"), seule une version du document étant prise en compte, les autres étant rejetées.

D'autres moteurs effectuent une sélection en fonction des formats de document (Excel, Powerpoint, PDF…) ou ils limitent leurs collectes à des fichiers d'une certaine taille (101 Ko pour Google - quoi qu'il semblerait que cette limite ait dernièrement "sauté" - 150 Ko pour MSN Search).

L'index

Le moteur d'indexation enrichit automatiquement un index des mots rencontrés. Cet index est constitué :

- D'un **index principal** ou **maître**, contenant l'ensemble du corpus de données capturé par le spider (URL et/ou document…).
- De **fichiers inverses** ou **index inversés**, créés autour de l'index principal et contenant tous les termes d'accès (mots clés) associés aux URL exactes des documents contenant ces termes sur le Web.

L'objectif des fichiers inverses est simple. Il s'agit de fichiers où sont répertoriés les différents termes rencontrés, chaque terme étant associé à toutes les pages où il figure. La recherche des documents dans lesquels ils sont présents s'en trouve ainsi fortement accélérée.
Comme pour les logiciels documentaires et les bases de données, une liste de mots "vides" (par exemple, "le", "la", "les", "et"…) est parfois automatiquement exclue (pour économiser de l'espace de stockage) ou ces mots sont systématiquement éliminés à l'occasion d'une requête (pour améliorer la rapidité des recherches).

Pour comprendre le fonctionnement d'un index inversé, prenons, par exemple, une page (http://xxx.sanglot.com/) comprenant la phrase "Les sanglots longs des violons de l'automne" et une autre page (http://yyy.violons.com/) contenant les mots "Les violons virtuoses ; les premiers violons du Philharmonique de Radio France". Les données suivantes figureront dans le fichier inverse :

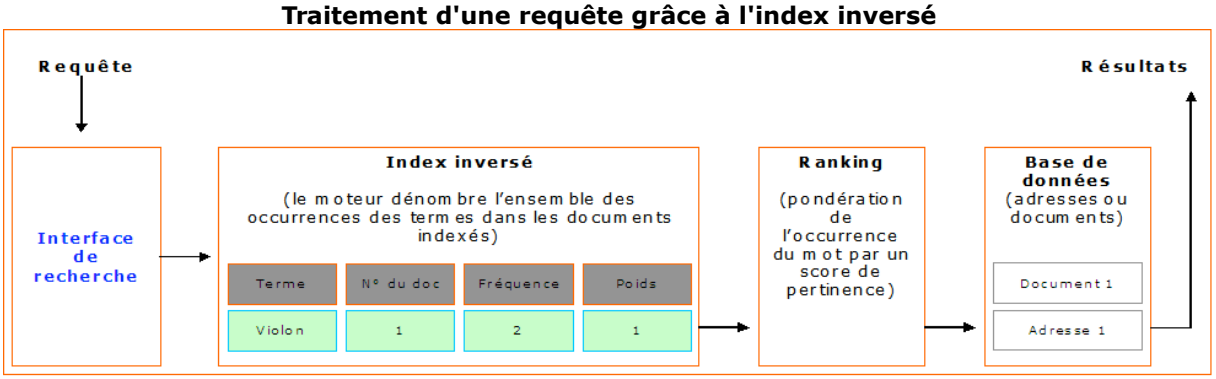
Exemple de fichier inverse

Terme	Numéro du doc. indexé	Fréquence	Poids			
			Titre	Adresse	Meta	Texte
Automne	1	1	-	-	-	1
France	2	1	-	-	-	1
Longs	1	1	-	-	-	1
Philharmonique	2	1	-	-	-	1
Premiers	2	1	-	-	-	1
Radio	2	1	-	-	-	1
Sanglots	1	2	-	1	-	1
Violons	1	1	-	-	-	1
	2	2	-	1	-	1

Virtuoses	1	1	-	-	-	1

Une requête dans le moteur de recherche avec le mot "violons" sera traitée en interrogeant l'index inversé pour dénombrer les occurrences de ce mot dans l'ensemble des documents indexés. Cette recherche donnera ici comme résultat les deux URL http://xxx.sanglot.com/ et http://yyy.violons.com/ et la page http://yyy.violons.com/ apparaîtra en premier dans la liste des résultats, le nombre d'occurrences du mot "violon" étant supérieur dans cette page. A noter toutefois, par rapport à cet exemple très "basique", que la fréquence des occurrences d'un mot sera pondérée par le processus de ranking des résultats (voir ci-après).

Traitement d'une requête grâce à l'index inversé



Note : Google associe également les termes des liens pointant vers une page (ancres) - autrement appelé "texte offshore des liens" - avec la page pointée (considérant que ces liens renvoyant vers une page fournissent souvent une description plus précise de la page que le document lui-même).

L'index doit être mis à jour régulièrement, en ajoutant, modifiant ou supprimant les différentes entrées. C'est en effet la fréquence de mise à jour d'un index qui fait en grande partie la qualité des résultats d'un moteur et sa valeur (pas de doublons ou de liens morts dans les résultats…), d'où des délais de rafraîchissement relativement courts.

Tailles de plusieurs index de moteurs

	Exalead	Gigablast	Google	MSN	Yahoo	Ask Jeeves
Nb de pages indexées	1 milliard	1,5 milliard	8 milliard	5 milliard	5 milliard	2,5 milliard
Source	Exalead	Gigablast	Google	MSN	SearchEngineWatch	
Date	02/2005	03/2005	11/2004	02/2005	11/2004	

Le système de ranking

Le ranking est un processus qui consiste pour le moteur à classer automatiquement les données de l'index, de façon à ce que, suite à une interrogation, les sites les plus importants, les plus pertinents, apparaissent en premier dans la liste de résultats. Le but du classement est d'afficher dans les 10 à 20 premières réponses les documents répondant le mieux à la question.

Les moteurs élaborent pour cela en permanence de nouveaux algorithmes (des formules mathématiques utilisées pour classer les documents). Ces algorithmes sont un véritable facteur différenciant. Ils ne sont donc que très rarement rendus publics et ils sont même dans certains cas protégés par des brevets.

Il existe trois grandes méthodes de ranking des résultats et les moteurs utilisent pour la plupart un mélange de ces différentes techniques (pour plus d'informations, voir l'article de Jean-Pierre LARDY sur les "Méthodes de tri des résultats des moteurs de recherche" : http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/00/53/sic_00000053_02/sic_00000053.html) :

- Le tri par pertinence

Les résultats d'une requête sont triés en fonction de six principaux facteurs appliqués aux termes de la question :

- "**Poids" et localisation d'un mot dans le document** (exemple : le poids est maximum si le mot apparaît dans le titre ou au début du texte) **ou son adresse** (url).
- **Densité d'un mot**, calculée en fonction de la fréquence d'occurrences du mot par rapport au nombre total de mots dans le document.
- **Mise en exergue d'un mot** : gras, titre (balise Hn), lien, italique, etc.
- **Poids d'un mot dans la base de données** calculé en fonction de la fréquence d'occurrence du mot dans l'index (les mots peu fréquents, rares, sont favorisés).
- **Correspondance d'expression** basée sur la similarité entre l'expression de la question et l'expression correspondante dans un document (un document est privilégie lorsqu'il contient une expression similaire à l'expression de la question).
- **Relation de proximité entre les termes de la question et les termes utilisés dans le document** (les termes sémantiquement proches sont favorisés).

- Le tri par popularité

Popularisé par Google en 1998 (pour contrer entre autres les abus possibles des méthodes de tri par pertinence), le tri par popularité s'appuie sur une méthode basée sur la co-citation et il est indépendant du contenu. Cette méthode de tri des résultats est aujourd'hui utilisée par de nombreux moteurs.

Google classe les documents en fonction de leur PageRank (nombres et qualité des liens pointant vers ces documents). Le moteur analyse en outre les pages contenant les liens (les liens des sites considérés comme "importants" pèsent plus "lourd" que les pages de certains forums jugés secondaires par exemple).

Presque tous les moteurs de recherche permettent de rechercher les liens pointant vers une page ou un site, appelés "backlinks" (en recherchant link:http://www.monsite.com).

- Le tri par mesure d'audience

Créée par la société DirectHit en 1998, cette méthode permet de trier les pages en fonction du nombre et de la "qualité" des visites qu'elles reçoivent. Le moteur analyse alors le comportement des internautes à chaque visite d'un lien depuis la page de résultats (et notament le fait qu'il revienne ou non sur le moteur et au bout de combien de temps) pour tenter de trouver les pages les plus "populaires" parmi les pages référencées et améliorer en conséquence leur classement dans les résultats. Cette méthode semble être tombée en désuétude depuis quelques temps.

- Le tri par catégories

Lancé en 1997, Northernlight proposait le classement automatique des documents trouvés dans des dossiers ou sous-dossiers (clustering) constitués en fonction des réponses. Les réponses intégrées à chaque dossier sont également triées par pertinence. Cette technique de "clusterisation" thématique des résultats est aujourd'hui notamment utilisée par le français Exalead et l'américain Vivisimo.

Les moteurs sont amenés à ajuster en permanence leurs algorithmes afin de contrer le "spamdexing", c'est-à-dire les techniques peu scrupuleuses de spamming utilisés par certains webmasters pour "tromper" les algorithmes des moteurs de recherche et améliorer artificiellement le positionnement d'une page.

Parmi les techniques les plus connues (et réprouvées par les moteurs), citons notamment le fait de multiplier les mots-clés dans les balises meta des pages HTML ; le fait d'intégrer un texte "invisible" sur une page (en blanc sur fond blanc, par exemple) ; la création de "sites miroirs" ou de liens fictifs ou invisibles pointant vers une page (ce qui permet de détourner l'indice de popularité) et, pour finir, la mise en place de faux portails contenant en fait des liens commerciaux ou le développement de "fermes de liens" (linkfarms), à savoir des listes de liens sans cohérence ayant pour unique objectif de gonfler la popularité des sites inscrits.

Le logiciel de recherche / moteur d'interrogation

Le moteur d'interrogation (*searcher*) est l'interface frontale proposée aux utilisateurs. Plusieurs niveaux de requête (simples ou avancées) sont en général offertes. A chaque question, par le biais d'un script CGI (*Common Gateway Interface*), une requête est générée dans la base de données et une page Web dynamique restitue les résultats généralement sous forme de liste ou de cartes de résultats. L'interface CGI permet d'exécuter un programme sur un serveur et de renvoyer le résultat à un navigateur Internet.

Focus sur le fonctionnement de Google

Architecture

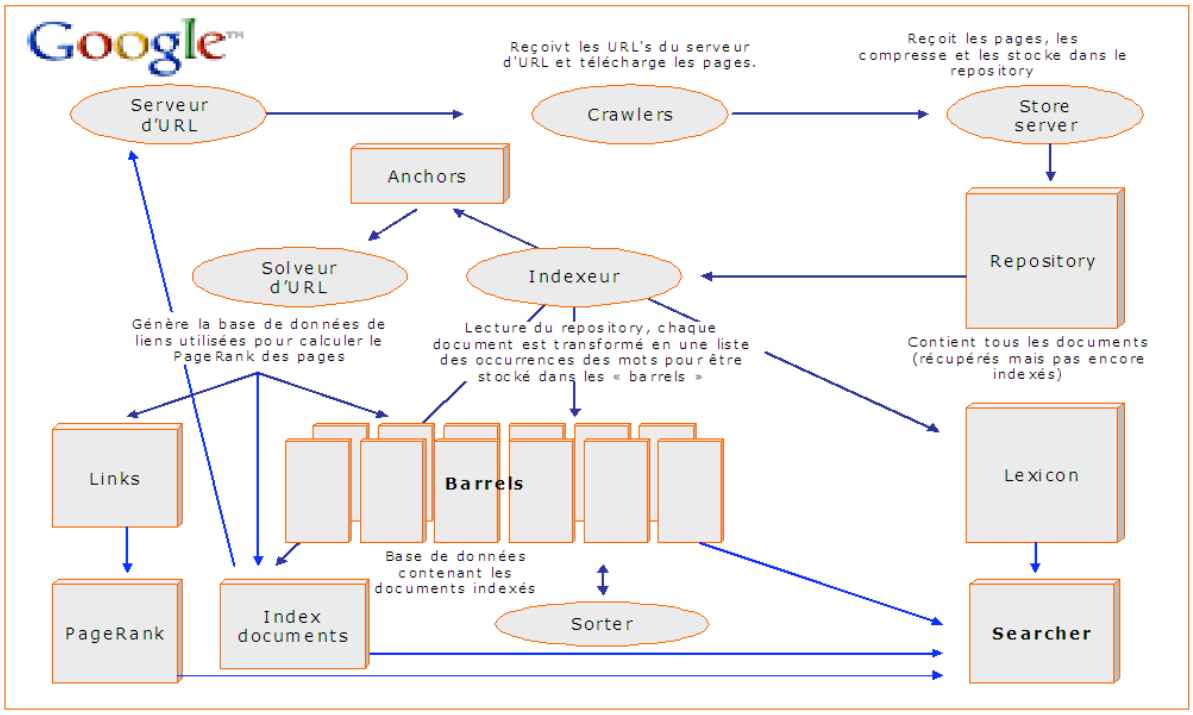
Créé en 1998 par deux étudiants de l'université de Stanford, Sergey Brin et Larry Page, Google s'est rapidement imposé comme le leader mondial des moteurs de recherche.

Le stockage des données et la réponse aux requêtes sont effectués à partir de dizaines de milliers de PC traditionnels tournant sous Linux. Réunis en clusters, les ordinateurs sont interconnectés entre eux selon un système basé sur la répartition des charges entre ordinateurs (un ordinateur distribue les tâches au fur et à mesure vers les autres ordinateurs disponibles).
D'un coût moins élevé que celui des serveurs, les PC traditionnels offre un avantage au moteur de recherche dans la mesure où il est possible d'agrandir relativement "facilement" le parc informatique à mesure que croissent le Web et la quantité de documents à indexer.

L'index de Google est découpé en petits segments (des "shards") afin qu'ils puissent être répartis sur l'ensemble des machines réparties dans des *data centers* déployés dans le monde entier, cela afin d'être toujours au plus proche des utilisateurs et de réduire au maximum les temps de réponse aux requêtes. Pour rester disponible en cas de défaillance d'un PC, chaque "shard" est dupliqué sur plusieurs machines. Plus le PageRank est élevé et plus le nombre de duplicata est élevé (voir http://www.webrankinfo.com/actualites/200411-infrastructure-google.htm).

Dévoilée il y a sept ans (et probablement toujours assez identique à l'heure actuelle) l'architecture de Google fait apparaître l'interconnexion de plusieurs composants séparés.

Architecture fonctionnelle de Google



Source : Sergey Brin et Lawrence Page - "The Anatomy of a Large-Scale Hypertextual Web Search Engine" (http://www-db.stanford.edu/~backrub/google.html)

Chaque composant a un rôle bien défini :

- Le **serveur d'URL** (URL server) envoie aux **crawlers** (Googlebot) toutes les adresses URL des pages devant être visitées (et notamment les liens soumis via le formulaire de soumission de Google http://www.google.fr/addurl/?hl=fr&continue=/addurl)

- Le **store server** compresse les pages extraites par les crawlers et les envoie au **Repository** où elles sont stockées

- L'**indexeur** lit et décompresse le contenu du Repository. Il associe chaque document avec un numéro identifiant docID et convertit chaque page en un ensemble d'occurrences de termes (chaque occurrence est appelé un "hit"), enregistrant les informations sur le "poids" du mot dans la page (position, taille de police…).

- L'**indexeur** distribue les occurrences dans un ensemble de "**barrels**" (organisé par docID).

- L'**Anchors** stocke certaines informations générées par l'indexeur, à savoir les liens hypertextes et les ancres qui leurs sont associés (textes des liens).

- Le **solveur d'URL** (URL Resolver) récupère les informations fournies par l'**Anchors** et convertit chaque adresse URL pointée par l'ancre en un docID (si cette adresse n'existe pas dans le Doc Index, alors il l'ajoute).

- Le **Links** contient des paires de docID (reçues du **solveur d'URL**). Il s'agit de paires de liens car chaque ancre appartient à une page et pointe vers une autre page.

- Le **PageRank** récupère les informations de cette base de données de liens pour calculer le PageRank de chaque document (indice de popularité).

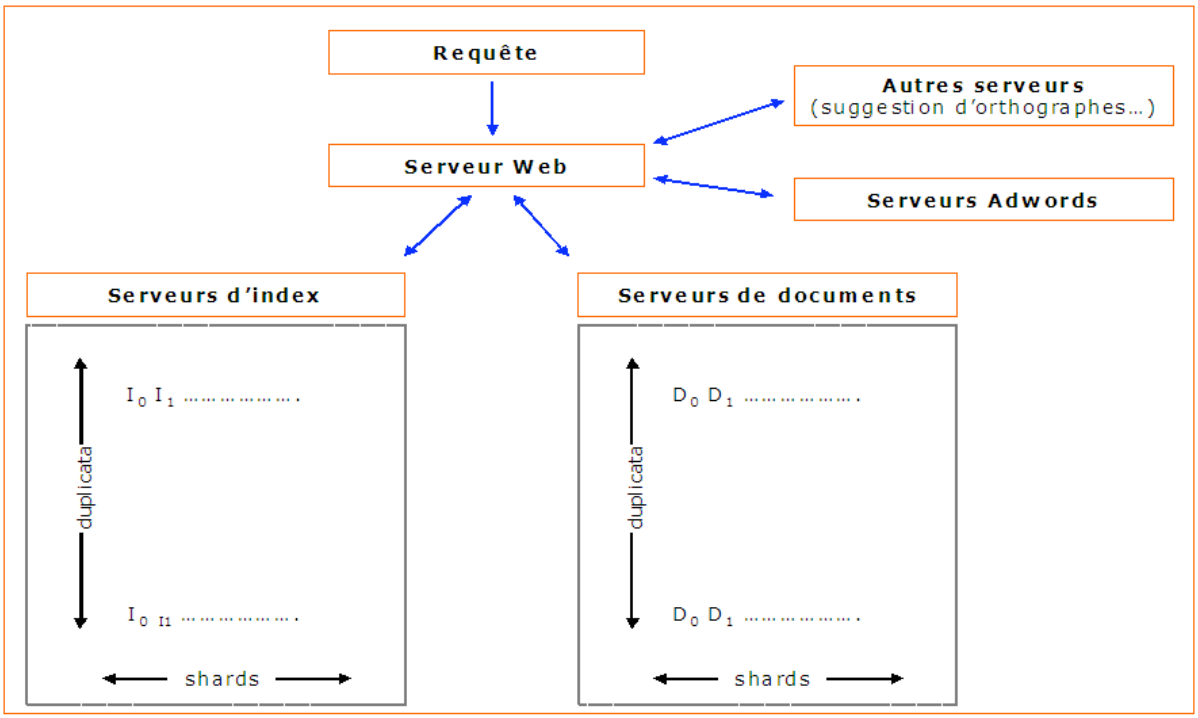
- Le **Sorter** récupère les données stockées dans les « **Barrels** », organisées par docID, et les réorganise en wordID (identités des mots). Cette opération permet de générer l'index inversé, stocké dans les même "Barrels".

- La liste des mots créée par le **Sorter** est comparée avec celle du **Lexicon** (lexique) et tout mot ne figurant pas dans le lexique y est ajouté.

- Enfin, le **Searcher** (interface de recherche) exécute les recherches pour répondre aux requêtes des utilisateurs. Il utilise pour cela le lexique (créé par l'indexeur), l'index inversé contenu dans les Barrels, les adresses URL associées aux mots de l'index inversé (provenant du Doc Index) et toutes les informations du PageRank concernant la popularité des pages.

A chaque requête, le serveur consulte l'index inversé et regroupe une liste de documents comprenant les termes de recherche (*hit list*). Il classe ensuite les pages en fonction d'indices de popularité et de pertinence.

Schéma de l'utilisation des serveurs de Google utilisés pour la réponse aux requêtes



Source : WebRankInfo (http://www.webrankinfo.com/actualites/200411-infrastructure-google.htm)

Crawling et indexation

Google se différencie des autres moteurs par la rapidité de son système de crawling.

Aujourd'hui quasiment disparue, la "Google Dance" désigne la période durant laquelle Google met à jour ses index (répartis dans un cluster de plusieurs dizaines de milliers de serveurs) et ses indices de pertinence / popularité. Une certaine instabilité dans les résultats témoignait alors du fait qu'une mise à jour de l'index principal était en cours (d'où le nom de "danse" de Google).
Cependant, depuis mi-2003, Google met son index à jour quasiment en "temps réel" et les mises à jour globales de l'index sont plus occasionnelles (pour plus d'informations, voir http://dance.efactory.de/ et http://www.webrankinfo.com/google/data-centers.php). La "Google Dance" désigne en fait aujourd'hui la période où Google met à jour le calcul du PageRank affiché dans sa barre d'outils et n'a plus aucune relation avec la notion de mise à jour de l'index en lui-même.

Ranking

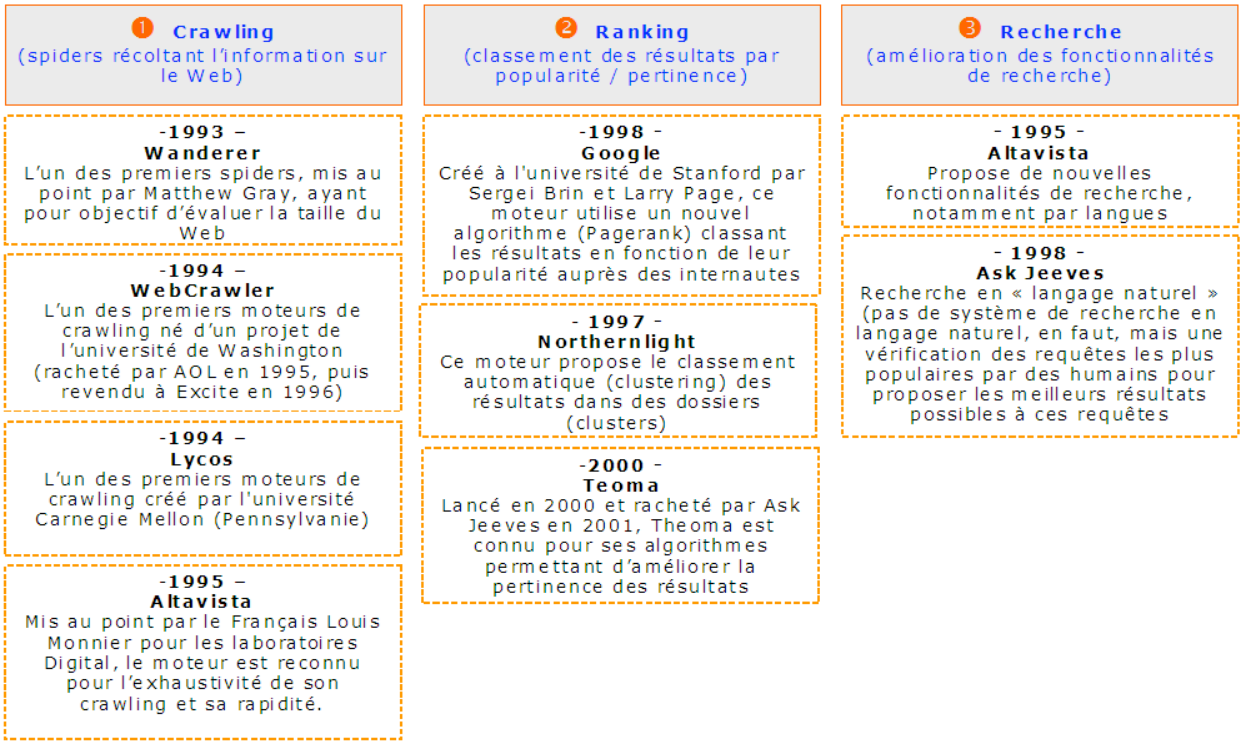
L'une des principales forces de Google tient, comme nous l'avons vu, à sa technologie de "PageRank", une technique qui permet d'appliquer la "notion de popularité" sur les sites indexés (plus il y a de page dans l'index du moteur qui proposent un lien vers le document à classer, et meilleur est le rang de celui-ci). Le champ d'évaluation de popularité s'étend également sur les pages pointant vers le document à classer (plus les sites qui pointent vers un site sont populaires et mieux les plages sont classées). Il s'agit ici de la notion d'"indice de popularité à deux niveaux".

Conclusion – Enjeux techniques pour les moteurs de recherche

Les moteurs de recherche arrivent désormais à un stade "mature" de leur développement. Pour la plupart, ils maîtrisent bien les grandes étapes technologiques de la recherche (crawling et indexation des pages, ranking et restitution des résultats). Selon François Bourdoncle, PDG d'Exalead, ils entrent donc progressivement dans une "logique produit" où les enjeux

technologiques se situent moins dans la capacité à traiter les informations en amont que dans l'ergonomie des interfaces et dans la qualité de restitution des résultats en aval (voir interview de à l'adresse http://www.01net.com/article/265903.html). Cela signifie-t-il que les moteurs de recherche ne peuvent faire mieux aujourd'hui en termes de pertinence ? L'avenir le dira, même si nous pensons, pour notre part, qu'il existe encore de nombreuses choses à faire dans ce domaine...

Bref historique des trois grands axes de développement technologique des moteurs



Source : SearchEngineWatch – 04/03/2004

Compte tenu de l'augmentation croissante du nombre de documents en ligne et de la diversification des formats indexés, les moteurs vont également devoir améliorer considérablement leurs capacités d'indexation et de rafraîchissement des contenus prochainement. Google vient notamment d'annoncer son intention de numériser des millions d'ouvrages provenant de bibliothèques américaines. Se pose la question du mode d'indexation qui sera retenu par le moteur pour ces documents ? S'agira-t-il d'indexation en texte intégral ?
Par ailleurs, le référencement de ces pages a priori "non Web" devra peut-être nécessiter une évolution du mode de ranking des résultats privilégié par Google. Le Pagerank, qui tient compte des liens hypertexte pointant vers une page pour déterminer son importance, ne pourra en effet s'appliquer aux ouvrages numérisés, qui ne contiennent pas de liens hypertextes.

Quelques liens...

Sur les robots
- http://www.robots.darkseoteam.com/ : L'observatoire des robots des moteurs de recherche Google, Yahoo et MSN (ce site fournit des statistiques de passage des robots Google, Yahoo et MSN sur sa page d'accueil).

- http://www.robotstxt.org/ : The Web Robots Pags (donne une liste des robots "actifs").

Sur Google
- http://www-db.stanford.edu/~backrub/google.html
(article des deux fondateurs de Google - Sergey Brin et Lawrence Page intitulé "The Anatomy of a Large-Scale Hypertextual Web Search Engine" et publié 1998).

- http://www.computer.org/micro/mi2003/m2022.pdf

(article de l'IEEE Computer Society – Web Search For A Planet : The Google Cluster Architecture).

Sur le PageRank
- http://www.pr10.fr/dernier-classement-complet-PR10.htm : PR 10
(site qui référence les pages en PageRank de 10 – PageRank le plus élevé - à chaque "Google Dance")

Lexiques sur les moteurs de recherche
- http://www.sumhit-referencement.com/savoir-lexique.asp
- http://www.dicodunet.com/definitions/moteurs-de-recherche/

Note : cet article a été écrit par un journaliste free-lance pour le compte d'Abondance