

LE PHENOMENE DE GOOGLELISATION DANS LES RECHERCHES D'INFORMATIONS SUR INTERNET.

Henry SAMIER*, Nicolas MOINET**, Pascal FRION***,
samier@istia.univ-angers.fr, nicolas.moinet@univ-poitiers.fr, pascal.frion@acrie.fr

* Enseignant chercheur, directeur ISTIA Innovation & Master Information Stratégique
ISTIA Innovation, Université d'Angers

62 avenue notre dame du lac, 49000 Angers – France, tél. : 33 (0) 2 41 22 65 49

** Maître de Conférences, directeur du Master Intelligence Economique & Communication Stratégique
ICOMTEC - Université de Poitiers - France

Chercheur au CEREGE - Equipe Intelligence Economique

<http://icomtec.univ-poitiers.fr>, <http://www.ie-poitiers.net>, tél. : 33 (0) 6 12 23 68 41

*** Animateur Acrie Réseau

Aéroport Nantes atlantique, rue Nungesser et Coli, 44860 Saint Aignan de Grand Lieu - France

Tél. : 33 (0) 2 40 04 25 25

Mots clés :

moteur de recherche - recherche d'information - surveillance de l'internet - collecte d'information

Keywords :

search engine - information retrieval - internet watch - information gathering -

Palabras clave :

motor de busqueda - busqueda de informacion - internet vigilancia - colecta de informacion

Résumé

Depuis ces dernières années, Google est devenu le moteur de recherche le plus utilisé sur Internet. Nous constatons que les internautes utilisent à quatre-vingt pour cent ce moteur en recherche simple et le questionnent avec deux voire trois mots clés. Cet usage simpliste et aveuglant induit une réduction du volume d'informations surveillées et un appauvrissement de l'analyse des données traitées. Nous proposons dans cet article d'analyser ce phénomène et de montrer comment dépasser cette googlelisation. Nous apportons une démarche méthodologique qui constitue une alternative à la logique du moteur unique et qui conduit d'une part à surveiller efficacement internet et d'autre part à gagner du temps dans les recherches d'informations existantes sur internet.

1 Les sept péchés capitaux de Google

Les outils de recherche sur Internet tels que les annuaires ont eu un développement logique dans les années 90, avec par exemple le succès de l'annuaire Yahoo. A cette époque les annuaires étaient majoritairement utilisés par les internautes. Les moteurs de recherche étaient moins utilisés comme par exemple Altavista et OpenText qui perdaient de l'argent. Les ressources informatiques nécessaires pour mettre en œuvre ces moteurs étaient importantes et les modèles économiques de « tiers payant ⁱ » étaient largement répandus. Pourquoi l'ascension du moteur de recherche Googleⁱⁱ, en à peine dix ans, a été aussi rapide et pourquoi l'usage de Google dépasse-t-il celui des annuaires et des autres moteurs de rechercheⁱⁱⁱ ?

Premièrement, Google a été reconnu pour être le plus rapide des moteurs de recherche. C'était le premier à répondre en moins d'une seconde à une requête tout en travaillant sur plusieurs milliards de pages. Cela a introduit un phénomène ergonomique et ludique dans les recherches d'informations où le plaisir individuel est devenu collectif par le bouche à oreille, dans l'échange d'adresses, la discussion sur des détails techniques et sur les actualités.

Deuxièmement, il a offert le plus grand volume d'informations accessibles avec un seul moteur. Ces deux faits ont une explication technologique. En effet, Google s'est fondé sur une architecture informatique distribuée en « cluster d'ordinateurs » qui a permis de fournir un temps de réponse très court tout en augmentant le volume d'information traité. Ainsi, Google utilise des milliers d'ordinateurs connectés en réseaux pour fournir le service de moteur de recherche. Les autres moteurs de recherche qui avaient une architecture informatique centralisée devaient augmenter la puissance de calcul en fonction du volume d'information pour garantir le même temps de réponse.

Troisièmement, Google s'est différencié par sa simplicité d'usage avec une page d'accueil "minimaliste". Voilà qui donnait une impression de pureté et d'objectivité et permettait un affichage rapide pour tous ceux qui ne sont pas connectés en haut débit. Aujourd'hui, cette interface est personnalisable et permet de densifier les informations choisies.

Quatrièmement, le moteur était accessible gratuitement 24h/24h et 7j/7J sans coupure de service (par exemple pour maintenance !) et était surtout en progression constante.

Cinquièmement, Google fut initialement l'un des premiers à fournir des résultats sans bandeau de publicité ni d'images "longues" à charger, mais il remplaça les bandeaux par des liens commerciaux contextuels. C'est-à-dire que les liens commerciaux proposés sont corrélés à la question ou aux mots clés saisis sur le moteur. Par ailleurs, nous constatons que les réponses du moteur sont majoritairement des réponses provenant de sites commerciaux.

Sixièmement, il offrait dans les résultats de nouvelles fonctions telles que « j'ai de la chance » et « pages similaires », des outils linguistiques tels que « traduire cette page » ou la possibilité de ne trouver que des documents en format pdf ou ppt. Aujourd'hui, la folle course au développement de nouvelles fonctionnalités est lancée avec, entre autres : la recherche par type de fichier, les images, la recherche multilingues, la recherche de ses propres documents sur son ordinateur, le emarketing avec Adwords, une barre de tâches, la recherche dans la presse, messagerie gratuite (Gmail), Google Groups (rachat de dejanews), la recherche dans les livres

(Google Book Search), recherche dans les brevets US (Google patents), Recherche dans les Blog (Google Blog), Moteur de bureau (Google desktop), Partage de documents (Google Doc, rachat de Writely et spreadsheet), La géolocalisation (Google Earth), ainsi que quelques autres services.

Septièmement, la méthode de classement était différente des autres moteurs. En effet, elle utilisait un nouvel algorithme de classement des résultats et se fondait sur la popularité des pages pour qu'elles soient en début de liste des résultats. Historiquement, la majorité des algorithmes des moteurs concurrents se fondaient sur la fréquence d'apparition des mots dans les pages. Aussi, l'esprit collaboratif tient au fait que Google ne pratique pas la vente de positions dans ses résultats, établissant son propre critère de pertinence « démocratique » avec le « PageRank ».

En résumé, les variables discriminantes de Google ont été la vitesse, le volume d'informations, la simplicité, l'accessibilité continue, l'absence de publicité non sollicitée, des nouvelles fonctions et la méthode de classement des résultats [8].

2 Le phénomène de googlelisation de l'Internet

Depuis sa création, Google n'a cessé d'augmenter son taux de pénétration sur le marché et décline ainsi ses concurrents. Par exemple pour la France, en 1998, il représentait 30% des usages alors qu'aujourd'hui il représente près de 90%^{iv}.

De nombreux auteurs ont fait le constat du danger de la googlelisation [2], [3], qui s'apparente à une forme de pensée unique et un phénomène de tropisme de la recherche d'information. Par exemple, PINTE [6] a révélé quelques dangers de la googlelisation dans l'enseignement et propose de dépasser les illusions de Google, par la préconisation d'une veille documentaire appliquée à l'université. Elle doit selon l'auteur se fonder sur une salle de veille pour favoriser le travail en groupe d'étudiants.

Google semble représenter aujourd'hui une sorte de lieu rassurant de par son caractère "unique". Ses contours semblent suffisamment larges et suffisamment établis pour que l'utilisateur n'ai pas l'impression de se perdre dans le « cyber-espace ». L'utilisateur voit Google comme un lieu unique, suffisant pour le temps à consacrer aux recherches.

L'internaute en vient à oublier les critères de recherche et les fonctionnalités qu'il souhaite dans sa démarche, car il s'en remet par défaut, et peut-être par dépit, à un outil unique. En 2004, dix outils de recherches gratuits parmi les plus populaires en France à cette époque, proposent plus d'une centaine de critères pour effectuer sa requête (voir illustration ci-dessous). Il résulte de ce recensement, que ces critères ne sont pas présents partout. Chaque outil propose ses propres critères et fonctionnalités. Certains critères ou fonctionnalités méthodologiques, pour une recherche d'information sont, quant à eux absents de chaque outil étudié (possibilité de réaliser un plan de recherche, mémorisation d'une requête, etc...).

Comment utiliser ? les techniques de recherche efficaces

- A quel moment faut-il utiliser des techniques ?
- Quelques astuces de recherche
- Les opérateurs logiques (booléens)
- La recherche avancée avec Altavista
- La recherche avancée avec Google
- La recherche avancée avec Voila
- Les fonctions de recherche classiques
- Le choix de l'outil pertinent pour ma recherche
- Résumé des possibilités de recherche dans un moteur
- De la recherche directe à la recherche indirecte

Testez-vous
Chapitre IV



Le choix de l'outil pertinent pour ma recherche

	Google.fr	Altavista.fr	Voila.fr	Tiscali.fr	Msn.fr	Exalead.fr	Lycos.fr	Yahoo.fr	Kartoo.fr	Alltheweb.com
Recherche de termes optionnels						✓				
Recherche par thématique			✓	✓			✓	✓	✓	
Recherche par public cible				✓						
Recherche par nature de la cible				✓						
Recherche par origine linguistique	✓	✓			✓	✓		✓	✓	✓
Recherche sur les racines des mots						✓				
Suggestion de mots clés ?	✓					✓				
Site sélectionné par Nomade				✓						
Critères temporels et géographiques										
Recherche par récence		✓				✓				
Recherche par mise à jour	✓			✓	✓	✓	✓	✓	✓	✓
Recherche par pertinence (ancien < > récent)						✓		✓		
Site récent ? (moins d'un mois)				✓						
Recherche géographique				✓						
Recherche restreinte au pays d'origine	✓	✓			✓	✓	✓	✓	✓	✓
Fonction "ne pas" (exclusion d'un pays)						✓				
Options des requêtes										
Plan de recherche										
Mémorisation d'une requête										
Nombre de requête mémorisables										
Taille de la requête mémorisable										
Opérateurs logiques										
Opérateur de présence obligatoire (AND)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Opérateur de présence facultative (OR)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Opérateur de refus (NOT)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Opérateur de proximité (NEAR)		✓	✓			✓	✓		✓	

Lors d'une recherche d'information sur internet, il conviendrait de préciser son projet et les données souhaitées, puis de choisir des critères de recherche et des fonctionnalités, afin d'identifier l'outil ou les outils pertinent(s) par rapports à nos attentes. Pour cela, il est nécessaire d'avoir une bonne connaissance manipulative des outils utilisés. Cette étude montre que les outils internet ne sont pas équivalents. Or les auteurs constatent que les internautes ont une faible connaissance de ces critères de recherches et des capacités des outils utilisés. Il n'y a pas de choix réfléchi d'un outil en fonction de sa recherche.

3 Comment sortir de la googlelisation : une méthode

Pour sortir d'une forme de « pensée unique » de Google nous préconisons l'usage successif des annuaires généralistes^v afin de préciser voir d'orienter les recherches, des encyclopédies et dictionnaires afin de compléter la liste des mots clés, des recherches d'expériences d'internautes telles que le « social search » et « social bookmark » afin de gagner du temps, des moteurs de recherche^{vi} et des méta-moteurs de recherche^{vii} pour prétendre à une plus grande exhaustivité.

Cette démarche est applicable dans les contextes scientifiques mais aussi dans les contextes courants de recherche d'information, de veille sur internet et plus généralement dans la démarche d'intelligence économique [4].

Plaçons-nous dans la situation d'un chef de projet ayant des recherches opérationnelles à réaliser. L'acquisition de données est vue ici comme un moyen, une action qui suit la formulation d'un projet et l'identification d'un manque et d'un besoin d'information et non pas un égarement dans l'inconnu pour se donner des idées.

La première étape consiste à définir le domaine de recherche, les mots clés et le thésaurus associés. Il est possible dans les cas de recherche exploratoire de nouveau domaine que la connaissance des mots clés soit limitée en début de recherche.

Dans ce cas nous préconisons l'usage :

- des annuaires généralistes afin de trouver les sites classés dans des catégories, et de trouver les catégories associées afin d'en extraire les mots clés qui seront utilisés dans les moteurs de recherche ;
- des annuaires de blogs^{viii} ;
- des dictionnaires et des encyclopédies^{ix} pour trouver les définitions, les experts, les synonymes et autres liens utiles ;
- des personnes sources, des accès spécialisés par domaine, comme par exemple les bases de données les brevets^x en ligne, ou encore les techniques de l'ingénieur.

Ainsi le chercheur a maintenant à sa disposition une première sélection de sites du domaine, des listes de mots clés et des expressions pertinentes.

Dans une deuxième étape nous préconisons l'usage des meta-moteurs afin sélectionner les moteurs pertinents pour le domaine de recherche concerné^{xi}.

La troisième étape est fondée sur la création de requêtes dans les moteurs de recherche, où nous préconisons l'usage de plusieurs moteurs afin de garantir un élargissement de la couverture du domaine. En effet, bien que Google soit un des plus volumineux moteur de recherche avec plus de 10 milliards de pages indexées, nous avons constaté qu'il n'y avait pas un total recouvrement du contenu de différents moteurs de recherche. Ainsi nous utiliserons en complément de Google, sept autres moteurs de recherche Web et des moteurs de recherche Web, Web 2.0 et Blogs^{xii}.

Par exemple, si le besoin de recherche d'informations porte sur une personne et ses réseaux associés, il est risqué d'utiliser uniquement Google [5]. Que ferez-vous face aux milliers de documents proposés ? Certes, les premiers documents nous donneront rapidement des renseignements mais après ? Allez-vous traiter l'ensemble des documents ? Allez-vous vous limiter aux trois premières pages au risque d'une vision partielle et partielle (les premiers résultats peuvent être le fruit d'un travail de référencement de la source elle-même) ?

En fonction de la requête vous aurez des résultats différents si vous saisissez : « prénom nom » ou « nom prénom » en phrase exacte ou encore « nom » ET prénom (en opérateur booléen). Si vous souhaitez aller vite et avoir une bonne vision de votre sujet, d'autres outils sont nécessaires. Par exemple, vous pourrez commencer par le métamoteur classique Metacrawler (www.metacrawler.com) et le métamoteur graphique Kartoo (www.kartoo.com) qui identifiera des mots clés et des sites ressources. Sur une carte seront visibles les liens existants entre les sites ainsi qu'une série de mots clés. Vous pourrez resserrer la recherche en éliminant les mots clés qui vous paraissent inadéquats ou en supprimant les sites non pertinents. Si votre « personne cible » travaille régulièrement avec les mêmes personnes, logiquement elles apparaîtront citées ensemble à une ou plusieurs reprises. Si votre cible dispose d'un site personnel pour se présenter, vous avez de bonnes chances de le trouver et de glaner des informations utiles.

Par exemple, si nous recherchons « Christian Harbulot », Directeur de l'Ecole de Guerre Economique, avec la démarche proposée et en quelques minutes il est possible d'avoir une première idée de son réseau. Pour aller plus loin, une fois découvert par exemple, que Christian Harbulot écrit des éditoriaux sur (www.infoguerre.com), vous pouvez essayer de cartographier les liens de ce site. Pour cela, nous pouvons utiliser entre autre Touchgraph (www.touchgraph.com) pour faire apparaître les liens similaires de google. C'est ensuite que nous pourrons utiliser Google en recherche avancée avec les fonctions du format de fichier, pour par exemple trouver des présentations powerpoint de Christian Harbulot. Sur « Google groupes », vous pourrez suivre certaines de ses interventions sur des forums ou repérer ses alliés et opposants. Avec les « alertes google », notamment programmées sur la fonction « Actualités », vous serez informé par mail des nouvelles informations disponibles et passerez ainsi de la recherche d'informations à la veille. Enfin, n'oubliez pas le web invisible, c'est-à-dire toutes les informations accessibles via Internet mais qui ne sont pas recensées par les moteurs de recherche (sans doute 80% du total). C'est par exemple le cas des archives des grands quotidiens. Vous voulez en savoir plus sur les réseaux de Christian Harbulot ? Allez par exemple sur www.latribune.fr (article « Patriote en col mao » d'Erwan Sez nec du 4 mars 2003) ou www.intelligenceonline.fr (archives et Top 100 de l'IE).

Enfin, il sera judicieux de présenter vos résultats sous forme de carte grâce aux outils de « mind mapping », un mode d'organisation des idées particulièrement efficace, qu'il s'agisse de prendre des notes, de conduire un projet, de prendre une décision ou de représenter un réseau. En effet, il est possible, de représenter de manière très lisible les différentes branches de ce réseau avec leurs membres, d'y insérer des liaisons entre membres, d'associer des symboles à chacun pour mieux le caractériser ou des liens hypertextes avec des sources externes (site internet) ou internes (rapport).

Pour avoir une vision panoramique des outils, il peut être utile de les situer par rapport au cycle du renseignement, fondement de tout système de veille [4]. Ainsi s'enchaînent les actions suivantes : trouver les sources pertinentes, les surveiller de manière systématique et intelligente, contribuer à analyser les données, les mettre en forme, cartographier leurs éventuelles relations, alerter les personnes concernées, stocker, diffuser et partager les informations.

La recherche d'information ne doit pas être une action distincte et finie avant la phase de traitement. Une démarche de recherche d'information doit être réalisée en plusieurs étapes, associant intimement du questionnement, de la recherche, du traitement, de la mise en forme des données acquises, puis, un nouveau questionnement et de nouvelles recherches. Une recherche collective sera souvent meilleure qu'une recherche individuelle.

4 Applications scientifiques

Dans les domaines scientifique il est plus pertinentes de réaliser dans une première étape des recherche sur des sources scientifiques qualifiées telles que des bases de données ou des sites web scientifiques. C'est seulement dans une seconde étape et fonction des besoins que l'utilisation des sources web, blog et wiki seront utilisées.

Parmi les bases de données nous trouvons principalement deux catégories, les sources classiques et les sources hybrides. Les bases de données dites « classiques » comme par exemple Dialog, Sciencedirect, STN, Questel-Orbit, Inspec, Rapra, et Metadex sont accessibles par abonnement et largement connu du monde scientifique. Les sources hybrides, aussi accessibles par abonnement, couplent des informations provenant de base de données et de web scientifiques. Par exemple Scopus (www.scopus.com) propose une interface avec un accès aux publications scientifiques, aux brevets et à une parti du web scientifique. Il est ainsi plus pertinent d'utiliser l'interface de scopus que celle de google dans n'importe qu'elle recherche d'informations scientifiques. Parmi les sources du web scientifique nous trouvons entre autre scirus.com et scholar.google.com mais nous

Par exemple si nous recherchons des informations suivantes :

Sur google pour les mots clés

- « social search » nous obtenons 1 610 000 documents, 17800 documents (méthode rapidfind [7]), 13200 documents (méthode linkfind [7])
- « social bookmark » nous obtenons : 31 800 000 documents, 122000 documents (méthode rapidfind), 41500 documents (méthode linkfind)
- « social bookmarking » nous obtenons : 11 300 000 documents, 1930000 documents (méthode rapidfind), 85100 documents (méthode linkfind)

Sur scholar.google.com pour les mots clés

- « social search » nous obtenons : 245 documents, 31 documents (méthode rapidfind)
- « social bookmark » nous obtenons : 221 documents, 3 documents (méthode rapidfind)
- « social bookmarking » nous obtenons : 1370 documents, 44 documents (méthode rapidfind)

Sur Scopus pour les mots clés

- « social search », nous obtenons : 6 publications scientifiques, 2086 documents « scientifiques » web et 16 brevets
- « social bookmark », nous obtenons : 4 publications scientifiques, 1139 documents « scientifiques » web et 2 brevets
- « social bookmarking », nous obtenons : 27 publications scientifiques, 52097 documents « scientifiques » web et 6 brevets

Sur Scirus pour les mots clés

- « social search » nous obtenons : 6 documents de journaux, 29 documents web scientifique et 2073 documents web
- « social bookmark » nous obtenons : 2 documents de journaux, 16 documents web scientifique et 1225 documents web
- « social bookmarking » nous obtenons : 11 documents de journaux, 62 documents web scientifique et 52041 documents web

Sur Del.icio.us

- « social search » nous obtenons : 4691 documents
- « social bookmark » nous obtenons : 3801 documents
- « social bookmarking » nous obtenons : 3584 documents

Sur Wasalive.com

- « social search » nous obtenons : 105 documents (depuis les 30 derniers jours)
- « social bookmark » nous obtenons : 17 documents (depuis les 30 derniers jours)
- « social bookmarking » : 145 nous obtenons : documents (depuis les 30 derniers jours)

Sur Rollyo.com

- « social search » nous obtenons : 1 110 000 documents
- « social bookmark » nous obtenons : 10 900 000 documents
- « social bookmarking » nous obtenons : 30 900 000 documents

Sur Blogdimension.com

- « social search » nous obtenons : 4691 documents
- « social bookmark » nous obtenons : 1791 documents
- « social bookmarking » nous obtenons : 10872 documents

Moteur	Mots clés	Social search	social bookmark	social bookmarking
	Gogle	1 610 000	31 800 000	11 300 000
	Scholar Google	245	221	1370
	Scopus	6 + 2086 + 16	4 + 1139 + 2	27 + 52097 + 6
	Scirus	6 + 29 + 2073	2 + 16 + 1225	11 + 62 + 52041
	Del.icio.us	4691	3801	3584
	Wasalive	105	17	145
	Rollyo	1 110 000	10 900 000	30 900 000
	Blogdimension	4691	1791	10872

Après analyse des 50 premiers résultats de chaque moteurs, nous avons évalué les taux de pertinence de chaque moteur :

Moteur	Google	Scholar Google	Scopus	Scirus	Del.icio.us	Wasalive	Rollyo	Blogdimension
pertinence	24 %	80 %	90 %	88 %	62 %	46 %	38 %	34 %

Il apparaît donc que Google a le moins bon taux de pertinence avec 24 % et au Scopus et Scirus ont les meilleurs taux avec respectivement 90 % et 88%. Au total sur les 236 documents pertinents il reste effectivement 143 documents pertinents après dédoublement.

5 Discussion

Nous avons montré au travers de ce cas que d'une part l'usage exclusif de google n'est pas satisfaisant et illustre le phénomène du « googlewashed » et d'autre part que le taux de pertinences des résultats de google est le plus faible. Nous voyons ici que la valeur d'une recherche résulte plus de la diversité des sources utilisées que d'un usage unique de moteur, qui se révèle pauvre, surtout quand l'outil tente d'orienter les résultats par des liens commerciaux. Le « syndrome du réverbère » (travailler sur les zones balisées et éviter les zones non éclairées) focalise le travail d'acquisition de données à la manipulation d'un outil « unique ».

Il apparaît une première confusion entre la recherche et la consultation d'internet, en effet une recherche pauvre dans sa requête s'apparente plus à une simple consultation de données existantes et formalisées sur internet. La seconde confusion se situe entre le besoin et le résultat c'est-à-dire que d'une côté le besoin porte sur un apport d'information pour prendre une décision et se positionner dans une situation délicate, alors que le résultat porte sur la consultation de ressources existantes à un instant (t) durant lequel nous allons nous en « servir ».

Selon nous il est utile de mettre en œuvre un apprentissage systématique de méthode de recherche auprès d'étudiants de Licence et Master et pas limiter cet apprentissage au seul usage des nouveaux outils ou à la simple techniques des « guillemets ». Nous constatons que généralement le temps de préparation d'une requête se limite souvent à quelques secondes voire, il peut être inexistant. Une étude [1] a été réalisée sur l'estimation du temps perdu dans les recherches d'informations avec les pourcentages suivants :

Pourcentage moyen de temps perdu sans formation aux méthodes			
Usage 1h/jour	2,5h/semaine	10h/mois	20h/an
35 %	50 %	60 %	70 %

L'auteur indique que sur un échantillon de 1500 personnes (cadres non cadres, jeunes ou plus âgés, étudiants en école de commerce et en école d'ingénieurs, du secteur public du secteur privé.) et sur la période [2000-2007], le gain de temps peut être estimé dans les proportions suivantes :

Pourcentage moyen de temps perdu sans formation aux méthodes			
Usage 1h/jour	2,5h/semaine	10h/mois	20h/an
25 %	35 %	40 %	45 %

Selon l'auteur il existe différents biais tels que les biais de prestige, cognitif, l'équivalence, de l'existant, volumique, analytique, de la réactivité positive et le biais de la ressemblance.

Le biais de prestige : "j'ai trouvé cette donnée sur Google, donc mon travail n'a pas à être remis en cause car Google est la référence mondiale" ;

Le biais cognitif de la première donnée nous dit que la première donnée que nous obtenons sur un sujet nouveau s'imprime dans notre cerveau plus vite et reste plus longtemps comme la référence inconsciente : bref, la première donnée, c'est la bonne !

Le biais de l'équivalence : une donnée en vaut une autre : "je n'ai pas besoin de regarder tous les résultats proposés par Google, je ne regarde que les résultats de la première page, et mon expérience me montre que les résultats dans les pages suivantes n'apportent pas un complément significatif" ;

Le biais de l'existant : l'analyse de l'existant est le modèle de base de l'analyse des données : "on va déjà regarder ce qui existe, ce sera déjà pas mal, vu la quantité de données disponibles que cela représente" ;

Le biais volumique : plus il y a de résultats proposés par un outil sur internet, mieux c'est ;

Le biais analytique : le traitement est mis en avant : filtrage, représentation graphique, statistique, etc.. ;

Le biais de la réactivité positive : nous sommes prêts à passer en revue pendant plusieurs minutes des résultats, du moment que nous n'avons pas perdu quelques secondes à préparer une requête de plusieurs mots et astérisques, parenthèses, opérateurs, fonctions spéciales, types de fichier désirés, dates voulues, langue, etc. ;

Le biais de la ressemblance : le système de référencement porte sur la base de la ressemblance (la convergence voire la consanguinité). Il est possible de demander un mot ou un ensemble de mots, mais il n'est pas possible de demander – par exemple – le contraire de ce mot ou de ce groupe de mots ;

6 Et les outils pour trouver ?

Existe-il des outils pour trouver ? L'outil génial est-il sorti ? Nous avons vu les risque que pouvait engendre un outil unique, donc notre quête portera sur le choix d'outil complémentaire d'aire à la recherche, au traitement et à la visualisation des résultats. Ainsi nous pouvons proposer l'usage des outils suivants :

L'exploration de l'encyclopédie wikipédia est améliorée par l'usage du site wikimindmap.org qui permet d'explorer graphiquement 7 sites de wikimédia. Dans cette même logique nous trouvons aussi le logiciel DeliciousMind [www.blainekendall.com/deliciousmind/], le logiciel Freemind [freemind.sourceforge.net/wiki/] et les sites linkedwords.com, mindomo.com, neuralmatters.com, visualthesaurus.com, yworks.com et visuwords.com. Pour d'autres outils de cartographie nous pouvons nous reporter entre autre sur le site mapdream.com

7 Conclusion

Nous avons vu que l'usage d'un outil unique tel que Google limitait l'étendue, la portée et la pertinence des recherches alors qu'avec une démarche logique qui utilise un panel d'outils nous augmentons le taux de couverture des recherches et le taux de pertinence des résultats. Pourtant le phénomène de googlelisation existe toujours et il est selon nous révélateur principalement d'un manque de méthode et d'une forme d'appauvrissement de la réflexion face à un écran. La qualité des résultats de recherche dépend de celle d'un raisonnement et d'une réflexion profonde sur le type de recherche plutôt que sur une logique d'outil. Les recherches d'information des internautes se fondent soit sur des choix d'investigation soit sur des choix d'ignorance. Aussi nous voyons non seulement que le manque de méthode est une perte de temps, mais aussi que l'usage de plusieurs outils se révèle indispensable et particulièrement les réseaux humains et sociaux [5]. Mais au-delà de ces considérations philosophiques se trouve une question plus complexe, à savoir la bonne compréhension de notions qui s'entremêlent : données – informations – connaissances.

Bibliographie

- [1] FRION P., Observaciones sobre la falta de método en la búsqueda de información en internet, Puzzle n°11, Barcelone. DL. N: B-45.448-03, Issn 1696-8573.
- [2] LAZULY P., Le Monde selon Google, Le Monde diplomatique, octobre 2003, [<http://www.monde-diplomatique.fr/2003/10/LAZULY/10471>]
- [3] LOCKERGNOME, Fighting the Googlization of Research, February 14, 2005 [<http://www.lockergnome.com/nexus/search/2005/02/14/fighting-the-googlization-of-research>]
- [4] MARCON C., MOINET N., L'intelligence économique, Editions Dunod, coll. Les Topos, 2006. ISBN : 9782100496341
- [5] MARCON C., MOINET N., Développez et activez vos réseaux relationnels, Editions Dunod, 2007 (2ème édition). ISBN : 9782100506743
- [6], PINTE J.P., La veille documentaire, une alternative à la googlelisation de l'information ?, SEMAFORAD-2, 2005, Université de Bejaia, et [veillepedagogique.blog.lemonde.fr].
- [7] SAMIER H., SANDOVAL V., La veille stratégique sur l'internet, Les éditions Hermès Sciences, Paris, 2002., ISBN : 2-7462-0325-1.
- [8] VISE D., MALSEED M., Google Story, Editions Dunod, 2006. ISBN : 2 10 049894 0

ⁱ Ce modèle permet aux utilisateurs de ne pas être les payeurs. Les payeurs sont ceux qui voient un intérêt à offrir des services dématérialisés aux utilisateurs.

ⁱⁱ Le nom de Google vient de l'unité mathématique googol qui est égale à 10^{100} .

ⁱⁱⁱ Se reporter par exemple à la page « Speed is of the essence » [<http://google.com/int/corporate/history.html>].

^{iv} Se reporter notamment aux sites [<http://barometre.secrets2moteurs.com>] et [www.xiti.com].

^v Par exemple yahoo.com (directory, guide web), indexa.fr, dmoz.com, galaxy.com/directory/, jayde.com, looksmart.com (directory), vlib.org, wedoo.com

^{vi} Par exemple altavista.com, alltheweb.com, antiseach.net, excite.com, exalead.com, go.com, hotbot.com, infotiger.com, lexee.com, live.com, lycos.com, mozbot.com, teoma.com, ujiko.com, webcrawler.com, webwombat.com, wotbox.com.

^{vii} Par exemple askjeeves.com, cometway.com, debriefing.com, dogpile.com, fazzle.com, ixquick.com, kartoo.com, mamma.com, metacrawler.com, profusion.com, scopsearch.com, thebighub.com, vivisimo.com, zapmeta.com

^{viii} Par exemple annublog.com, blogarama.com, bloogz.com, bloguniverse.com, bloghub.com, go2web20.net, lamooche.com, rssreporter.net, retronimo.com, scienceport.org.

^{ix} Par exemple wikipedia.com

^x Par exemple plutarque.com, patft.uspto.gov (ou google.com/advanced_patent_search), ep.espacenet.com, delphion.com, freepatentsonline.com et patents1.ic.gc.ca/srch_adv-e.html

^{xi} Par exemple ariane6.com, clusty.com, datavor.com, dogpile.com, fazzle.com, fooxx.com, ithaki.net, ixquick.com, jux2.com, kartoo.com, killerinfo.com, mamma.com, metacrawler.com, search.com, searchbug.com, scopsearch.com, sidekiq.com, thebighub.com, veoda.com, vivisimo.com, vroosh.com, zapmeta.com

^{xii} Par exemple : Wasalive.com, Simpy.com, Oskope.com, reddit.com, Chacha.com, Yopple.net, Gnosh.org, Quintura.com, Tafiti.com, Cuill.com, Ziki, Blogdimension, technorati.com, blogsearch.google.com, feedster.com, blogpulse.com, blogsearchengine.com, bloggernity.com, blogdigger.com, search.blogger.com, blogolist.com, completerss.com, rssfeeds.com, popdex.com, blawgs.detod.com, boogieplay.com, search4blogs.com, blogdump.com, ask.com, feedburner.com