



Les journées d'études

L'avenir de la recherche d'information sur le web

"Peut-on vivre sans Google ?"

Mardi 11 octobre 2005

L'avenir de la recherche d'information sur le web

"Peut-on vivre sans Google ?"

**Le mardi 11 octobre 2005 de 9h30-17h15
à l'ENS Lettres et Sciences Humaines – Lyon**

9h00/9h30 - Présentation et Animation

Olivier Andrieu Abondance.com et co-créateur du moteur de recherche Mozbot.fr

9h45/10h45 - Où va Google ? A quoi ressemblera ce moteur dans deux ans ? L'hégémonie de Google représente-t-elle un risque ? Les moteurs de recherche, levier stratégique du 21ème siècle : pourquoi l'Europe est-elle absente de ce marché ?

Jérôme Charron, [Motrech](#) Blog et liste de diffusion

Frank Poisson, [consultant et éditeur de Blog à Part](#) (ancien Directeur Commercial de Google France)

Jean Veronis, [CILSH](#), [et Aixtal blogspot](#)

11h15/12h15 - Desktop Search : lier Web mondial et système d'exploitation local : faut-il éviter les outils de Google et MSN ? Les outils de Desktop Search sont-ils dangereux pour la vie privée ? La recherche sur les PC (et sur le Web en général) peut-elle être envahie par le marché publicitaire ?

Sébastien Billard, [blog Référencement, design et cie](#)

Olivier Ertzscheid, [Université de Nantes](#) et [Affordance.info](#)

Frank Poisson, [consultant et éditeur de Blog à Part](#) (ancien Directeur Commercial de Google France)

14h15/15h15 - Personnalisation, contextualisation, statistique, sémantique, géolocalisation, moteurs thématiques et régionaux : quoi de neuf en 2005 ? Quelles ont été les innovations principales en 2005 en termes de recherche d'information ?

Jérôme Charron, [Motrech](#) Blog et liste de diffusion

Fabrice Lacroix, [Antidot](#)

Gilles Serrafero, [Indexa](#)

Armelle Thomas, [Inforizon](#)

14h15/15h15 - L'avenir de la recherche d'information : vers où s'oriente-t-on ? Tentative de prospective... Qui pour concurrencer Google dans les années qui viennent ? La concurrence à Google viendra-t-elle de l'Open Source ? A quoi ressemblera le moteur de recherche "idéal" demain

Sébastien Billard, [blog Référencement, design et cie](#)

Jérôme Charron, [Motrech](#) Blog et liste de diffusion

Armelle Thomas, [Inforizon](#)

Jean Veronis, [CILSH](#), [et Aixtal blogspot](#)

15h45/16h45 - Conclusion

Olivier Andrieu Abondance.com et co-créateur du moteur de recherche Mozbot.fr

Le paysage des moteurs de recherche en 2005

Parts de trafic des différents moteurs

- 1. Google 70 à 80%
-
- 2. Yahoo 6 à 9%
-
- 3. Wanadoo/Voila 3 à 9%
-
- 4. MSN 4 à 5%
-
- 5. AOL 2 à 4%
-
- 6. Free 1 à 2%
-
- 7. Lycos, AltaVista, Tiscali, Club-Internet : 0,4 à 0,7%

Suivent :

AllTheWeb, Virgilio, Netscape, Mozbot...

Résultats par technologie :

- 1. Google 80%
-
- 2. Yahoo 10%
-
- 3. Wanadoo 3 à 9%
-
- 4. MSN 4 à 5%
-
- 5. Exalead 2 à 4%

Où en est-on fin 2005 ?

Déclin des annuaires généralistes.

L'immense majorité du trafic est générée par des : 1) moteurs 2) généralistes.

99% du trafic est généré par moins de 8 outils et 5 technologies.

Peut-on vivre sans Google

11 octobre 2005

Il existe de moins en moins de technologies différentes.

La plupart des technologies clonent Google.

Yahoo! et MSN n'ont pas, pour l'instant, réellement transformé leur essai. Google, plus que jamais est en position d'hégémonie.

Bref, il existe un vrai risque de « monolithisation » des moteurs, dangereuse pour tous !

Olivier Ertzscheid

[Université de Nantes](#) et [Affordance.info](#)

Entre monomanie ...



& mégalomanie ...

La question posée maintenant est celle des risques présents dans le croisement du web mondial et de la recherche locale, sur nos ordinateurs portables.

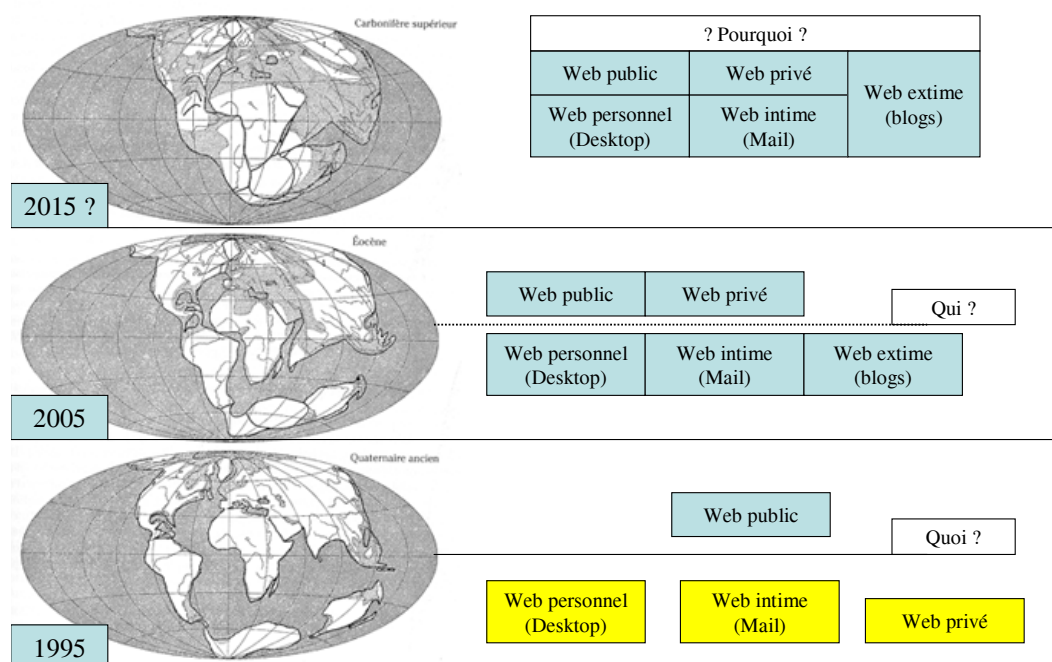
Les mêmes outils, le moteur Google est l'un d'entre eux, permettent aujourd'hui de chercher indistinctement dans la toile mondiale et dans les tréfonds de nos disques durs.

Essayons de retracer rapidement l'évolution qui nous mène à cette situation avant d'en tirer des constats et d'en analyser les risques ou les opportunités.

CHAPITRE PREMIER : LA DERIVE DES CONTINENTS

Dérive des continents informationnels.

indexé
 non-indexé



Le web naît officiellement en 1994 avec la publication d'un article de Tim Berners Lee « L'hypertexte et le CERN ». Pour que le web existe il faut la conjonction de trois éléments distincts : des adresses (URL) permettant de localiser l'information, des navigateurs (permettant d'y accéder) et un format d'encodage (HTML) permettant d'afficher l'information récupérée. Sans oublier les protocoles d'échange de données (http et autres TCP-IP). Petit à petit cette formidable base de connaissance et d'information qu'est le web prends son « corps digital. »

Vers la fin des années 90 (**première période**) nous disposons ainsi d'un web public (le « www ») indexé par les moteurs et contenant différents types d'informations. A ses côtés, un web privé, opaque, se constitue via les sites d'entreprise (notamment) qui déploient derrière leurs pages web des bases de données d'informations dynamiques (php entre autres), invisibles à l'œil des moteurs. Tout cela cohabite avec nos habitudes de travail et le contenu de nos disques durs, « évidemment » inaccessible aux moteurs ainsi qu'avec un web « intime » celui constitué de la trame

Peut-on vivre sans Google
11 octobre 2005

7/37

de nos messages électroniques (mail). Des ces 4 briques (web public, privé, disques durs, web intime), seule la première est accessible aux moteurs.

La question qui permet alors de scinder l'information en « visible / invisible », « indexée / non-indexée » est celle de la nature de l'information : Quoi indexer ? A laquelle on répond logiquement : « ce qui est délibérément mis en accès dans les espaces indexables du web public. »

Un pas est franchi (**deuxième période**) à l'heure actuelle avec « l'indexabilité » des 4 espaces décrits ci-dessus. Il n'est plus aucun d'entre eux pour échapper à ces formidables machines à indexer que sont les moteurs, Google en tête : le web public se déploie de manière exponentielle, le web privé ou invisible l'est de moins en moins (le premier symptôme fut l'indexation des fichiers « non-web » de type .doc ou .pdf) et dans ce même espace les entreprises emploient in interne des solutions de recherche « corporate » vendues par les mêmes acteurs (là encore, mais il n'est pas le seul, Google), mais surtout, le contenu de nos mails bascule vers des services de webmail avec des espaces de stockage défiant l'imagination (Mailnation.net offre gratuitement 1000 Gigas de stockage ...), et last but not least, nos disques durs eux-mêmes sont désormais indexés et visibles par les moteurs. Il n'est donc a priori plus rien de notre vie intime et professionnelle qui échappe (théoriquement) à ces étranges scrutateurs. Cette part d'intimité dévoilée, indexée, est même accrue par le phénomène des blogs qui déversent sur la toile les états d'âme d'une multitude d'individualités.

La question clé pour saisir ce qui se joue à l'heure actuelle n'est donc plus de savoir « quoi indexer ? », tout étant indexable, mais de savoir « qui indexe », ou plus précisément « qui a capacité de détenir des information suffisamment complètes sur nos vies professionnelles et intimes pour pouvoir à partir de là regrouper des bases de données hétérogènes et les utiliser à des fins commerciales ou autres. » Mais là encore la réponse est déjà posée. Google mais aussi Yahoo ! et MSN disposent d'outils et de services web qui leur permettent cela. Pour en rester sur le cas Google, Gmail collecte et scanne le contenu de nos mails, Blogger indexe nos états d'âme de blogueurs, Gdesktop vient fouiner dans nos disques durs.

Pour l'instant (et pour ce que l'on en sait) ces bases de données, ces bases d'index sont encore distinctes et liées au service qui les supporte. Mais la société mère est la même. C'est Google qui est le dépositaire de l'ensemble de ces données librement ou distraitemment confiées.

Et nous voilà à la **troisième période** de cette dérive des continents informationnels : la fusion complète et totale des services et des bases de données associées au sein d'un seul et même index. La question posée est ici celle que John Battelle appelle « la base de données des intentions ». Après le « quoi indexer » de la première période, après le « qui indexe quoi » de la seconde, il faut maintenant déterminer le « Pour quoi cette indexation ? » Et ici la réponse s'affirme comme une évidence : il s'agit de monétiser, de marchandiser des services, des biens, des échanges, des

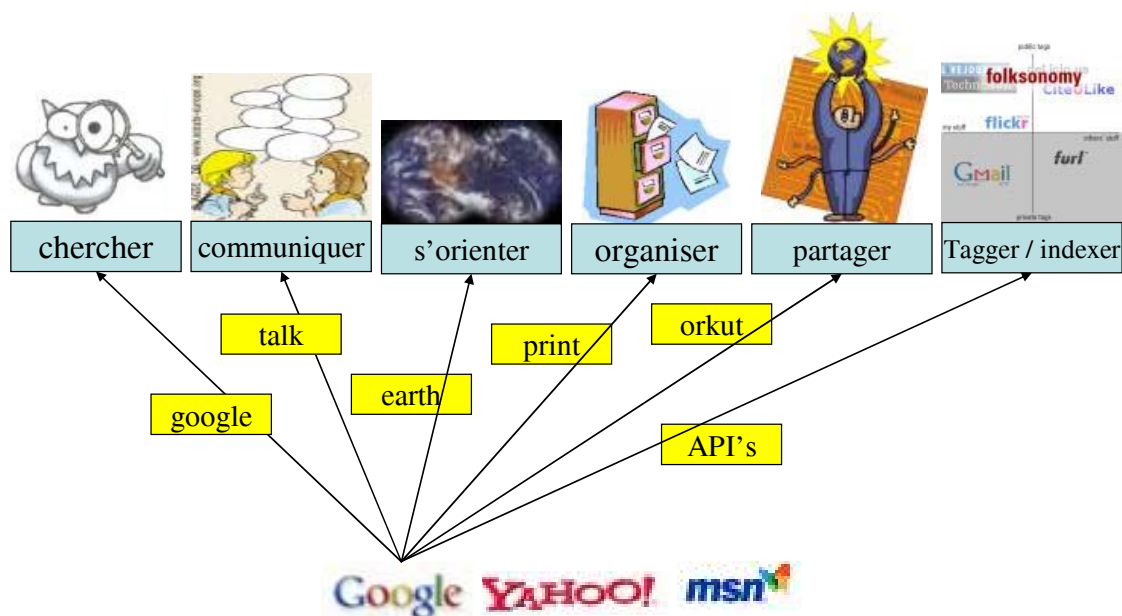
idées ... Car de l'interopérabilité entre ces différentes bases de données, Google a dès aujourd'hui la possibilité d'établir des inférences et des recoupements pour savoir qui nous voyons, ce que nous nous disons dans nos mails, avec qui nous échangeons, sur quels sites naviguons nous le plus souvent, quels livres ou quels types de biens nous achetons en ligne. Bref un ensemble d'éléments qui valent de l'or quand il s'agit d'établir des campagnes de publicité ciblées ou d'afficher des publicités contextuelles lorsque nous relevons notre courrier électronique.

Donc résumons : **la dérive des continents informationnels décrit le mouvement de ces gisements d'information qui, à l'inverse de la marche du monde, se rapprochent et fusionnent inexorablement entre les mains d'un très petit nombre de sociétés commerciales dont l'objectif avoué est de maintenir en vie un modèle économique consistant à monétiser et à marchandiser le maximum de services et de biens communs d'information possibles.**

On l'aura compris, cette dérive n'est pas sans conséquences (doux euphémisme) sur notre manière d'interagir en réseau et sur nos modes de traitement, de recherche et d'accès à l'information. Elle impacte directement notre environnement cognitif : témoin cette stupeur lorsque j'ai vu pour la première fois s'afficher sur la page de résultats de google (je venais d'installer l'outil de desktop searching) des pages web « classiques » et des documents dormant dans les entrailles de mon disque dur.

CHAPITRE SECOND : LA CONFUSION DES PRATIQUES INFORMATIONNELLES

Confusion des pratiques informationnelles



L'un des corrolaires directs de cette étrange hybridation entre recherche locale et globale est d'entraîner une grande confusion des pratiques informationnelles. Surtout qu'à l'heure actuelle ces dernières sont loin de se résumer à une activité de « recherche » ou « d'information retrieval » comme disent les anglo-saxons : « chercher » donc mais aussi « communiquer », « partager », « annoter », « organiser », « s'orienter » sont autant de verbes d'actions engagés dans nos pratiques quotidiennes d'information. Et là encore, Google et ses deux acolytes (Yahoo ! et MSN) brouillent considérablement les cartes.

Voici une petite liste de pratiques informationnelles et des outils associés. Pour le seul Google il nous est aujourd'hui possible de :

- chercher (via le moteur Google) dans des espaces, on l'a vu auparavant, distincts.
- Communiquer de personne à personne via Google Talk
- S'orienter à l'échelle du monde (via Google Earth) ou à celle de notre quartier (via Google Local et Google Maps)
- Partager des contacts et des réseaux sociaux via Orkut
- Diffuser de l'information (intime ou non) via Blogger

- Tagger/indexer des pages web (sur le principe des folksonomies) via différentes API disponibles autour de Google
- et l'on pourrait continuer la liste ad aeternam : acheter, vendre, comparer (Froogle), correspondre (Gmail) ...

Sans oublier la plus problématique de ces pratiques informationnelles, la plus spécifique parce que la plus liée aux capacités du raisonnement humain : ORGANISER. Or il s'agit, on le sait de l'ambition et de la devise affichée de Google : « *Organiser l'information à l'échelle de la planète.* » Admettons (par manque de temps) qu'il s'agisse là d'une compétence entrant dans les capacités d'un outil reposant presque exclusivement sur des logiques statistiques de matching, et qu'il ne soit besoin d'aucune capacité intellectuelle particulière pour mettre en place une bibliothèque mondiale (Google Print) ... admettons disais-je même si l'on aura compris l'inanité de la précédente proposition ...

Or bien plus qu'une organisation des connaissances (« informations » serait déjà plus réaliste) à l'échelle de la planète c'est à une véritable dérégulation des pratiques que nous convient nos chers moteurs au travers des services (certes parfois innovants) qu'ils mettent en place et surtout au travers de la fusion de ces services entre eux au nom d'une personnalisation toujours accrue (l'exemple en la matière étant donné par Yahoo ! et l'ensemble des services regroupés sous MyWeb2.0 ou encore Yahoo!360)

Envisageons le problème selon deux axes : en bas, sur l'axe des abscisses (PERMETTRE DE), la longue liste des pratiques informationnelles précitées (« *rechercher, organiser, communiquer, échanger, tagger* » ...). A la verticale (VENDRE) , sur l'axe des ordonnées, la non moins longue liste des services marchands (ou en passe de le devenir) proposés par les moteurs : il s'agit de vendre des livres, des accès Wifi, de la téléphonie, des mots-clés payants, demain peut-être des laptop ou des environnements web pour remplacer les OS existants ...

La question qui se pose pour les trois grands (GYM : Google, Yahoo ! MSN) ou plus exactement la question que nous devons leur poser sachant que c'est l'ensemble des croisements possibles entre les deux axes qui règle leurs logiques commerciales, la question que nous devons leur poser, disais-je, est celle de savoir :

- Selon quels référents et avec quelles limites ces croisements seront faits ?
- Quels sont les modèles sur lesquels ils comptent s'appuyer pour mener à bien certains de ces croisements ? Songeons à la bibliothèque universelle selon Google (Print) peut-on mener au sein d'un même espace, d'une même interface une mission de mise à disposition des connaissances pour le plus grand nombre et une autre de rentabilité pour la vente d'ouvrages ? Le modèle préconisé par les adeptes de bibliothéconomie n'est pas exactement le même que celui préconisé par les intérêts boursiers de Wall Street ...

Peut-on vivre sans Google

11 octobre 2005

11/37

- Quelle est la logique de ces sociétés : « donner accès » ou « prendre (des) parts » ??
- Enfin quels croisements seront faits entre quels services pour éviter la mise en place de ces « *bases de données des intentions* » qui mettent en jeu et en cause notre droit à « s'informer en paix » sans avoir à chaque instant de questions à se poser sur « comment et par qui seront réutilisées les informations que je mets en ligne. »

Alors peut-on vivre sans Google ? Bien évidemment.

Peut-on s'informer sans Google ? Aujourd'hui oui. Mais demain ?

Comme l'a sûrement rappelé Jean Véronis, les pires scénarios Orwelliens sont imaginables. Plus nous serons à en débattre, plus nous serons à affirmer la nécessité absolue d'une vraie pédagogie de l'information (et des outils d'information) à l'école et à l'université, plus nous serons à réfléchir ensemble aux alternatives possibles en évitant de sombrer dans un discours catastrophiste oublieux des extraordinaires avancées que ces outils ont permis pour la démocratisation et l'accès à l'information, et plus lesdits scénarios orwelliens auront de chance de se dissiper.



Pass the Cheetos Stop Bogarting Look at the Colors

What was I saying?

Advanced Toking
Mumbling Tools

Thought Search

I'm Feeling Stoned

Une chose m'importe encore : rappeler qu'aujourd'hui quand nous évoquons la question des outils de recherche (Google ou un autre mais Google en particulier) nous ne devons pas nous tromper d'échelle : nous ne parlons pas d'informatique et il ne s'agit plus de technologie. Et il n'est plus acceptable d'affirmer (comme le fit encore il y a peu Eric Schmitt, CEO de Google interrogé sur la part de Google dans la censure des sites accessibles depuis la chine) que « les technologies sont neutres ». Les technologies ne valent que par leur USAGE : l'histoire y compris la plus tragique nous l'a suffisamment montré. A nous (et à d'autres) d'allier pédagogie et réflexion politique sur ce que nous voulons faire demain d'un monde d'informations pour une économie de la connaissance.

Le chapitre troisième reste donc à écrire. Au delà des percées technologiques qui guideront ses pas (web sémantique) il ne pourra pas faire l'économie d'une réflexion politique sur les enjeux de la maîtrise individuelle de l'information pour la co-construction d'une connaissance collective.

Sébastien Billard, [blog Référencement, design et cie](#)

Les bonnes questions à se poser :

Ne sommes nous pas dépendants de Google ?

Ne passons nous pas à coté de ressources interessantes ?

L'utilisation de certains outils présente-t-elle des dangers ?

Quel futur pour la recherche d'information ?

Existe-t-il d'autres façons de chercher ?

Desktop Search

Ces outils répondent-ils à un besoin

Ces applications sont tout simplement des outils de recherche locale. Ils ne sont pas nouveaux, et les moteurs ne les ont pas inventé. Mais c'est seulement aujourd'hui qu'ils deviennent grand public et que les moteurs s'y intéressent (certains se contentant juste d'apposer leur marque sur des produits préexistants comme Yahoo/X1).

Tous les grands moteurs ont leur outil de recherche locale : Google DS, Yahoo DS, barre MSN DS, Askjeeves DS, Exalead One

Pourquoi cette eclosion soudaine : Il va de plus en plus y avoir une demande : Internet, le haut débit, le multimedia ont contribué à multiplier le nombre de données stockées et échangées sur nos ordinateurs. Nos données ont donc tendance à "enfler", à la fois en complexité et en volume, à un tel point qu'il n'est parfois plus possible de les organiser dans une hiérarchie. C'est un peu la même chose qui c'est passé pour le web : les annuaire à structure hiérarchique ont laissé la place aux moteurs algorithmique, plus à même de chercher dans de très larges collections de documents hétérogènes (Ce qui ne veut pas dire que les annuaires ont perdu de leur intérêt).

Mais ne nous leurrions pas : c'est avant tout pour capter l'utilisateur afin de générer du revenu...

Web et OS sont-ils véritablement liés :

Oui et non. Non parce que web et données locales restent bien séparés : les données locales ne restent (heureusement) accessible que localement. Par contre oui, au niveau des recherches de l'utilisateur, et au niveau de l'interface, la frontière se fait plus floue, une recherche portant à la fois sur les données distantes et locales.

Faut-il avoir peur de ces outils ?

Dangers potentiels :

On pense tout de suite à Big Brother... Scenario peu probable mais profiling publicitaire éventuellement possible. On peut compter cependant sur la vigilance de la communauté (googlewatch, blogosphère...) : il est relativement facile de repérer qu'un logiciel envoie des données.

Certains logiciels insistent sur le fait qu'ils ne se connectent au web que sur demande, comme Exalead One qui vise le marché professionnel.

Dangers réels :

Menaces au niveau local : le principal danger ne vient pas de l'extérieur mais de l'intérieur : vos données, l'historique de vos recherches, vos conversations, le contenu de vos emails sont stockés localement et donc accessibles par d'autres utilisateurs du poste. Il arrive que des documents détruits soient toujours présents dans le cache de l'outil de Desktop Search, ce qui donne accès à de l'information qui devrait normalement ne plus exister... Risque donc de vol d'information, ce qui peut être un problème en entreprise, quand on est un chercheur, ou même au sein du couple ;)

Menaces venant du réseau : Les outils de Desktop Search centralisant vos informations, il est possible, voir probable qu'ils soient la cible de hackers (Google DS a déjà corrigé des failles).

La publicité dans ces outils

La publicité contextuelle est actuellement absente des outils, car risque de suspicion évident : comment le moteur afficherait-t-il des annonces ciblées si les données locales n'étaient pas analysées à distance ?

Mais aucune donnée n'est elle réellement transmise ? Quid de la connexion avec les autres services des moteurs (Google personalized, Orkut, identifiant unique) ?

Avenir de la recherche

Vers où s'orientent-ils ? Pas de direction unique, mais des pistes.

Peut-on vivre sans Google

11 octobre 2005

15/37

L'évolution se fait à 2 niveaux : ce qu'on cherche (texte, multimedia, blogs...), et comment on le cherche.

Moteurs spécialisés :

A la différence des moteurs généralistes, les moteurs de recherche spécialisés ont vocation à répondre à une recherche particulière, mais à le faire très bien. Ce qui est spécifique dans l'utilisation de ces moteurs n'est donc pas la méthode de recherche, mais surtout l'objet de la recherche (Avec bien sûr des algorithmes spécifiques au type de contenu cherché).

On a ainsi vu fleurir des moteurs spécialisés pour chaque type de contenu ou de recherche : Technorati, Blogdigger, Feedster et autres pour les blogs et fils RSS ; Blinkx, Truveo pour la recherche vidéo et/ou audio ; Answers.com pour la recherche de réponses.

Recherche personnalisée :

La recherche personnalisée se base sur le fait que l'individu est unique, et la notion de pertinence subjective. On va donc permettre à l'individu de personnaliser sa recherche (dans une certaine mesure). La recherche personnalisée donne plus de contrôle à la personne : on peut filtrer pour exclure des sites considéré comme indésirables (Google Personalized Search, Yahoo My Web 2.0, Mozbot, Ujiko), ou au contraire filtrer pour inclure uniquement certaines sources (Rollyo, qui permet la recherche sur une sélection restreinte de sites).

Mais danger d'atteinte à la vie privée (Google traque les urls visitées, les recherches sont spécifiquement liées à un utilisateur identifié), et de perte de contrôle : la personnalisation affectant les résultats, il faut être conscient que nos recherches et personnalisations précédentes affectent les pages de résultat actuelles.

Moteurs à curseur :

Ici, l'individu interagit avec le moteur en fonction de ce qu'il juge pertinent sur le moment, en faisant varier certains paramètres comme la fraîcheur des documents, leur popularité... Des exemples réussis (Yahoo mindset, MSN Search Builder), d'autres moins à mon sens (Grokker, Kartoo), car interfaces à l'usabilité médiocre. Il ne faut pas oublier que l'individu généralement aime la simplicité (par ex, les booléens sont peu utilisés alors qu'il existent depuis le début des moteurs), les réponses immédiates (Jakob Nielsen réaffirmait récemment la prédominance des paramètres par défaut). Mais pour les "travailleurs du savoir" ou les chercheurs motivés, ce sont des fonctionnalités très intéressantes.

Moteurs de croisement ou à folksonomie :

Avec ce type de moteur, ce que le moteur analyse n'est pas tant la ressource elle-même que ce que disent les individus à propos de la ressource.

L'individu est donc la référence, et non le contenu : les individus forment des réseaux sociaux, qui catégorisent l'information, la filtrent naturellement (pas de spam), et dont l'étude du comportement sert à caractériser les ressources pertinentes. Exemple : Yoono, une application de partage de favoris. Ou Yahoo myweb 2.0.

Technologies du langage :

Les moteurs vont probablement de plus en plus intégrer une analyse sémantique des contenus. Une des applications actuelles est par exemple la clusterisation d'Exalead/Aol.fr : les résultats de recherche sont accompagnés de dossiers thématiques permettant d'affiner la recherche. (Parole au spécialiste Jean Veronis).

Qui pour concurrencer Google ?

De nombreux moteurs disparus ces dernières années (Altavista, Fast, Inktomi, Northern Light...), on aurait pu croire à un moteur unique, mais on assiste à un renouveau de la recherche. 4-5 grands moteurs aujourd'hui à la pertinence satisfaisante : La guerre de la pertinence est-elle terminée, laissant la place à une guerre des fonctionnalités ? (Opinion de François Bourdoncle, d'Exalead).

Google reste actuellement la référence. Mais en le considérant ainsi ne passe-t-on pas à côté du nouveau Google ? Si Google est une référence en général, de nombreux services spécialisés sont bien plus performants.

L'open source : intéressant sur des projets spécifiques, des doutes pour un grand moteur mondial (besoins de gros moyens).

Ne va-t-on pas également moins rechercher, ou du moins moins activement ? Remarque de Peter Da Vanzo de Search Engine Blog, qui cherche moins et qui trouve de plus en plus de ressources web via des applications de partage de favoris ou des blogs, alors qu'il ne cherchait pas particulièrement. Pouvoir des recommandations, du marketing viral : le contenu pertinent vient à nous sans que nous le cherchions.

Url citées

<http://desktop.google.fr>
<http://desktop.yahoo.com/>
<http://toolbar.msn.fr>
<http://sp.ask.com/docs/desktop/>
<http://beta.exalead.fr/search>
<http://www.google-watch.org/>
http://www.securinfo.umontreal.ca/ressources/indexation_poste/
<http://www.technorati.com/>
<http://www.blogdigger.com/>
<http://www.feedster.com/>
<http://www.blinkx.tv/>
<http://www.truveo.com/>
<http://www.answers.com/>
<http://www.google.com/ig>
<http://myweb2.search.yahoo.com/>
<http://mozbot.fr/>
<http://www.ujiko.com>
<http://mindset.research.yahoo.com/>
<http://search.msn.com/>
<http://www.rollyo.com/>
<http://www.grokker.com>
<http://www.kartoo.fr/>
<http://www.yoono.com/>
<http://www.useit.com/alertbox/defaults.html>

Gilles Serrafero, [Indexa](#)



l'avenir de la recherche d'information sur le web
« Peut-on vivre sans Google ? »

(Journée du 11 Octobre 2005)

Recherche d'information sur le Web

- complémentarité entre annuaire & moteur de recherche
- navigabilité sémantique par catégorisation
- fiabilité et mise à jour de l'information
- corrélation des réponses obtenues
- descriptif standardisé
- extension à un contenu partenaire suivant le contexte trouvé

(moteur) (filtres) (annuaire)

www.indexa.fr

Présentation d'un annuaire

(Étude : INDEXA.fr)

- point d'entrée par l'annuaire ou le moteur de recherche interne
- organisation en multi-annuaires :
 - *filtres* : géolocalisation / vente en ligne / fiches entreprises
- contenu propriétaire et indépendant suivant une ligne éditoriale
- rattachement d'un contenu partenaire :
 - *liens* : Bilan (societe.com) / Emploi : (emailjob.com) / ...

(géolocalisation) (liens) (rubriques)

http://www.indexa.fr
L'annuaire Internet des professionnels



l'avenir de la recherche d'information sur le web
« Peut-on vivre sans Google ? »

(Journée du 11 Octobre 2005)

Constat

- mise à jour du contenu lente & fastidieuse
- coût important de maintenance

Conséquences

- disparition des annuaires généralistes
- émergence de petits annuaires mutualisés
- apparition de répertoires de liens sponsorisés
- arrivée de pseudo-annuaires (farm link)

Solutions

- externaliser les mises à jour (fiches entreprises)
- faire un moteur qui indexe les pages des sites d'un annuaire :

fiche de DocForum



http://www.indexa.fr
L'annuaire Internet des professionnels



Peut-on vivre sans Google
11 octobre 2005



Principe et avantages de la solution AFS

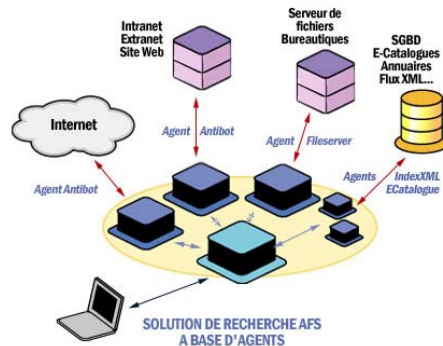
AFS est une solution de recherche d'informations multi-sources à base d'agents logiciels répartis spécialisés dans des domaines de compétence.

Une stratégie unique qui offre de nombreux avantages :

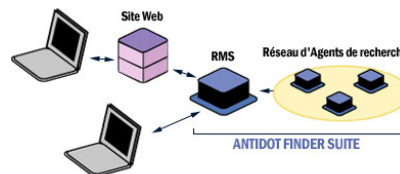
- ✓ **maximise la pertinence** car chaque agent exploite au mieux les spécificités des sources de données
- ✓ permet d'intégrer naturellement les **nouveaux formats** et les nouveaux supports
- ✓ garantit la **pérennité** du système : évolutivité, scalabilité
- ✓ permet un **déploiement simple** grâce à une collection d'agents génériques prêts à l'emploi.

Intégration rapide, facile et souple :

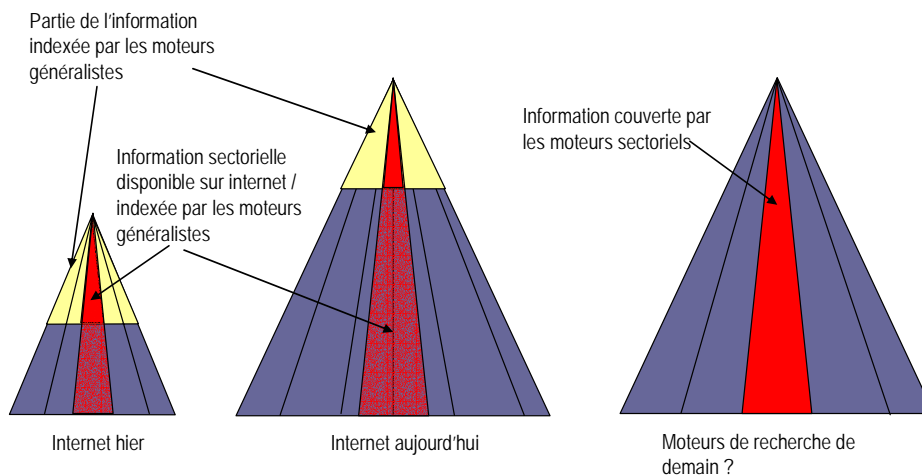
- ✓ le moteur est accessible comme Web Service
- ✓ les résultats sont nativement en XML donc grâce à l'emploi de feuilles de styles (XSL, CSS) la présentation est complètement paramétrable et facile à faire évoluer



Une mécanique de propagation / agrégation de flux XML avec des règles externes de traitement des flux.



Moteurs de recherche Internet



Les moteurs de recherche généralistes ont une vision de plus en plus limitée de l'info disponible.

Les moteurs sectoriels permettent de couvrir un sujet de façon exhaustive.

Moteurs sectoriels

- ✓ **Couverture maximale de l'information**
 - indexation en profondeur de l'information disponible sur internet
 - possibilité d'ajouter des bases de données ou de connaissances sectorielles

- ✓ **Homogénéité de l'information indexée**
 - catégorisation de l'information en sujets cohérents
 - suggestion de documents connexes

- ✓ **Exploitation classique et variée**
 - moteur de recherche
 - surveillance de contenus (veille, plagiat)

- ✓ **Exemples :** www.batiactu.com recherche.assurance-maladie.fr

[batiactu.com]



5

Moteurs thématiques

- ✓ **Sortir de l'approche documentaire**
 - passer du document à « l'objet »
 - rendre les objets vivants afin de les faire interagir

- ✓ **Mettre en œuvre conjointement**
 - **des technologies de textmining**
extraction d'entités nommées, de concepts, ...
 - **les technologies du Web Sémantique**
représentation de la connaissance : ontologies, TopicMaps

- ✓ **Illustration par le contre-exemple**
 - www.froogle.com : comparateur de prix
 - news.google.com : recherche de news

6

Moteurs thématiques

✓ Cas du Tourisme

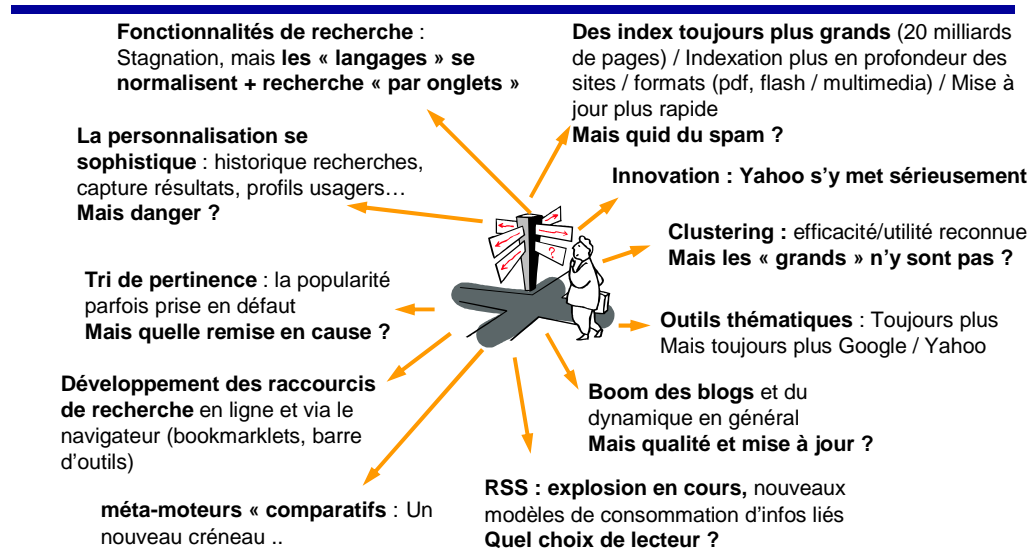
- Indexation de sites Internet et consolidation avec des bases de données qualifiées
- Transformer les documents en objets touristiques vivants : hôtels, gîtes, spectacles, terrain de sport, musée, ...
- Faire interagir les objets entre eux afin de les combiner et de proposer aux utilisateurs des réponses riches à des recherches complexes (composition de séjours, ...)

✓ **Produit ITis (co-édité avec la société Mondéca) : 3 ans de développement (projet européen Craft), partenariats avec des labos.**

7

« Entre les outils d'alerte et les résultats RSS, je me rends compte je travaille davantage sur des sortes de mers intérieures plutôt que sur le grand magma du Web »

2005 : Confirmation de tendances plus que nouveautés réelles



©Armelle Thomas - Octobre 2005

2

Quel moteur de recherche « idéal » pour demain ? Quelques (modestes) pistes...

Un moteur qui cherche à « comprendre » en amont le besoin de l'internaute (piste Yahoo!Mindset) sans vouloir trop en savoir sur lui, sauf exception...

Un moteur qui met plus d'intelligence dans ses rouages (notamment linguistique)

Un moteur qui donne automatiquement de bons points de départ (cf Teoma)



Un moteur qui « exploite » les avancées du web sémantique (pourquoi pas des recherches par type d'acteur, par type de sources)

Un moteur qui propose différents accès à l'information (dans différents « fonds », recherche intégrale et classification, en donnant le choix sur le critère de pertinence à privilégier,...)

©Armelle Thomas - Octobre 2005

3

Olivier Andrieu, [Abondance](#)

« Je suis passé par tous les états dans cette journée.

Je me demandais ce matin s'il fallait continuer de faire des recherches avec Google, eu égard à son cote Big Brother et cette après-midi, on a assisté à un foisonnement d'idées pour améliorer les outils et l'on a vu qu'en terme d'hégémonie, rien n'est joué.

Il y a 13 ans que je suis dans la recherche d'information et je suis sûr que l'on n'en est encore aujourd'hui qu'à 10% de ce que l'on peut faire. »

Olivier Andrieu constate qu'il y a un chaînon manquant entre le cerveau de l'utilisateur et l'expression de la requête.

C'est très dur de comprendre ce que l'utilisateur a cherché en voyant les mots-clés. Même moi, quand je cherche des informations, je m'y reprends à plusieurs fois pour reformuler la requête »

Armelle Thomas pense qu'il faut enseigner la recherche d'information. Une matière à débat pour Olivier Andrieu, qui se demande si ce n'est pas l'outil qui doit s'adapter à l'utilisateur ou l'inverse.

Mais une « mauvaise » formulation peut aboutir à des résultats très éloignés de ce que l'on recherche, Jean Véronis le démontre en citant des exemples de recherches banales sur Google, pour lesquelles s'affichent des liens sponsorisés plus que tendancieux....

A son avis, le problème du langage a été sous-estimé.

La catégorisation mériterait d'être affinée et il faudrait essayer de « désambigüiser » les termes et les expressions.

Il faudrait travailler sur le multilinguisme, car il ne faut pas oublier que 35% des internautes ne parlent pas anglais.

La traduction automatique, est loin d'avoir atteint le meilleur, pouvoir lire les résultats dans une autre langue que celle dans laquelle ils ont été créés sera une véritable avancée.

Le moteur idéal de demain ?

L'avenir

Une chose m'a beaucoup fait sourire à la fin de cette journée: Sans même nous être concertés, en utilisant des chemins et une argumentation totalement différentes, Jean et moi même en sommes venus à la même conclusion, la seule entité, aujourd'hui capable de concurrencer Google est bel et bien Microsoft. Au delà du futur concurrent de Google, l'avenir semble se dessiner selon cinq axes essentiels:

- L'utilisation de la linguistique dans les processus d'indexation et de recherche. Les moteurs généralistes que nous connaissons aujourd'hui utilisent des technologies des années 70! Je le soulignais déjà en 1998 dans "Évaluation/Comparaison des outils de recherche sur Internet", depuis, rien n'a changé, et Jean et Armelle se sont fait un devoir de nous le rappeler.

- L'amélioration de la présentation des résultats. Pourquoi toujours des clones de Google? Pourquoi les interfaces ne sont-elles pas plus visuelles? (J'ai toujours été étonné de ne pas voir une plus grande utilisation du FishEye).
- L'amélioration de l'utilisation du clustering. En effet, les algorithmes de clustering sont aujourd'hui bien connus, mais il faudrait étendre leur utilisation. Comme l'a souligné Jean à plusieurs reprises, il est incroyable qu'un de ses billets soit le premier sur Google sur la recherche "vernis à ongles". Il suggère donc d'être en mesure de calculer le score d'une page, non seulement en fonction du contenu de celle-ci (et de ses liens), mais également en fonction de l'adéquation qu'il existe entre la question et l'ensemble du site auquel appartient une page. C'est en effet une très bonne idée qui améliorerait certainement beaucoup les résultats. Reste dans un premier temps à définir la notion de site... car si celle de la page n'est déjà pas si évidente qu'il n'y paraît, celle du site me semble encore bien plus complexe (quoi qu'il en soit, c'est tout de même une excellente idée qu'il conviendra de creuser dans Nutch).
- La lutte contre le spam.
- La prise en compte des méta-données afin de mieux qualifier l'information (attention au spam) : Description, keywords, dublin-core, micro-formats, ...

Et le lendemain du 11 octobre on en parlait sur les blogs en direct

Sébastien Billard, [blog Référencement, design et cie](#),
le 12/10/2005 :

Organisée par Docforum, [cette table ronde](#) a été l'occasion de réfléchir ensemble sur le phénomène Google, ses dangers et dérives éventuelles, ainsi que sur les nouvelles tendances en matière de recherche d'information.

Très rapide condensé de ce qui s'est dit :

[Jean Veronis](#), tout à la fois pertinent et impertinent, nous dépeint un Google "dealer", qui veut nous rendre dépendants, et s'inquiète de la prédominance américaine dans l'accès à l'information : quel accès européens ou français avons-nous ? Voila est moribond, et l'on ne sait pas trop ce que fait Exalead, bien que celui-ci soit très pertinent... Jean nous rappelle également que la recherche, c'est avant tout du langage, et que les moteurs sont à la traîne dans ce domaine, utilisant des méthodes élaborées dans les années 70. Une plus grande utilisation des technologies du langage est donc une piste à explorer, qui pourrait passer par une analyse sémantique globale des sites, le multilinguisme... Jean pense d'ailleurs que Microsoft a les moyens scientifiques et bien sûr financiers de produire des outils de recherche efficaces.

Pour [Jérôme Charron](#), il ne faut plus attendre d'évolutions majeures du moteur, mais seulement des services connexes. Google veut nous rendre accros, et le profiling quasi-systématique l'inquiète. Il est de plus en plus difficile d'arbitrer entre services personnalisés et vie privée. Jérôme déplore le manque de réelles innovations de la part des moteurs, tout en voyant de grandes possibilités dans l'utilisation des [folksonomies](#), un point de vue généralement partagé par tous les intervenants. Selon [Olivier Andrieu](#) Yahoo travaillerait d'ailleurs activement sur le domaine.

[Franck Poisson](#), considère que le terme "moteur de recherche" est dépassé, et lui préfère le terme d'"IUI", ou "Interface Utilisateur d'Information". L'avenir de Google est un "webtop", sorte de plateforme ou desktop basé sur le web. L'interface aura de plus en plus tendance à s'adapter aux recherches de l'utilisateur. Selon lui, les parts de marché des différents moteurs ne devraient pas évoluer sensiblement dans les 4 à 5 années à venir.

Peut-on vivre sans Google
11 octobre 2005

26/37

[Armelle Thomas](#) souligne que le critère de popularité (PageRank) est parfois pris en défaut, mais quelle alternative à celui-ci ? Armelle note également que la recherche "hors web", notamment via RSS est amenée à se développer, point de vue que je partage totalement.

[Fabrice Lacroix](#) plaide pour les moteurs thématiques et sectoriels, permettant une meilleure couverture sur un domaine donné que les moteurs de recherche généralistes.

Jérôme Charron, [Motrech](#),

le 12/10/2005 :

Mardi dernier se tenaient à l'ENS-Ish une journée organisée par l'association DocForum sur le thème "Peut-on vivre sans Google?". Fabuleuse journée

Occasion de rencontrer hors de la matrice des gens passionnants: L'animateur Olivier Andrieu (il aura fallu presque dix ans pour que nos premiers échanges électroniques se matérialisent dans la vraie vie), les intervenants, Franck Poisson, Jean Véronis, Sébastien Billard, Fabrice Lacroix, Gilles Serraféro, Armelle Thomas, mais également une assistance nombreuse, intéressée et très réactive.

"Occasion de réfléchir ensemble sur le phénomène Google, ses dangers et dérives éventuelles, ainsi que sur les nouvelles tendances en matière de recherche d'information" (Sébastien Billard).

Beaucoup de choses très intéressantes ont été évoquées. Ceci n'est absolument pas un compte-rendu, mais plutôt les quelques notes qu'il me reste en tête après avoir laissé reposer un peu mes idées. Je regroupe ici les différentes idées échangées selon les cinq thèmes essentiels qui ont émergés lors de cette journée:

Le business,

la confidentialité,

les moteurs thématiques,

le Web 2.0

l'avenir.

Business

Franck (que je me félicite d'avoir suggéré comme intervenant, cette journée sans lui aurait été très différente) nous rappelle à juste titre que Google c'est avant tout une machine à dollars (oui, je sais, je suis trop idéaliste pour accepter cette idée). Et Fabrice de reprendre un peu plus tard: D'un point de vue financier, pourquoi Google aurait-il intérêt à ajouter de nouvelles fonctionnalités à son moteur? En effet, lorsqu'on domine plus de 80% du marché, pourquoi innover? Lorsque, la pompe à \$ est en marche, et débite à plein régime, pourquoi prendre le risque de faire baisser son débit? La situation quasi monopolistique de Google l'empêche d'innover, puisque le coût d'intégration d'une nouvelle fonctionnalité à son moteur ne pourra jamais être couverte par les revenus supplémentaires générés par cette nouvelle fonctionnalité. Exemple: Imaginons que Google (avec 80% de part de marché) génère 80\$/an par le biais de son moteur, et que l'intégration d'une nouvelle fonctionnalité lui coûte 2\$/an. Il faudrait alors pour que cet investissement soit rentable que les revenus du moteur dépassent à 82\$/an, c'est à dire que sa part de marché dépasse 82%. Or il est très difficile de gagner 2 points de part de marché, lorsqu'on en possède déjà 80. Bilan: Pourquoi, tant qu'il n'y a pas de vraie concurrence, apporter de nouvelles fonctionnalités? (C'est exactement la même situation que celle d'Internet Explorer avant que Firefox ne vienne changer la donne). Bref, le monopole tue l'innovation, car, rappelez vous, tout ça n'est finalement que du business, rien de plus. A propos de business, et en regardant de plus près celui de Google, nous sommes arrivés à la conclusion que le modèle économique de Google est finalement extrêmement fragile. Il ne repose que sur l'affichage de publicités adaptées au contenu (AdSense) ou de mots achetés (AdWords). Bref, des systèmes potentiellement faillibles, sur lesquels les annonceurs ont peu de contrôle: Google exploite l'ignorance de ses annonceurs.

Confidentialité et vie privée

Par le biais d'approches très diverses, nous étions tous à peu près d'accord sur le risque de l'hégémonie de Google. L'accent a été mis sur les dangers d'un accès à l'information sur Internet dominé par l'industrie américaine, mais également sur les risques concernant la protection de la vie privée. L'exemple donné par Franck est saisissant: Ouvrez un email sur Gmail. Regardez maintenant à droite de cet email. Que voyez-vous? Des pubs. Ok, il faut bien financer le service. Mais rien ne vous choque? Regardez bien. Oui, tout à fait. Les pubs sont en rapport avec le sujet du mail. Cela ne vous choque pas plus que ça. Bon. Maintenant, sortez de la matrice pour quelques instants. Revenez dans la vraie vie, et imaginez. Imaginez que votre facteur ouvre votre courrier, le lise, et glisse dans chaque enveloppe des publicités en rapport avec son contenu. Cela vous parle un peu plus? Cela vous choque? Vous trouvez ça inacceptable? Et bien c'est ce que fait Google avec vos emails.....

Moteurs thématiques et sectoriels

Fabrice a présenté ce qu'étaient les moteurs thématiques. Une excellente présentation qui a achevé de me convaincre de leur utilité et de leur avenir très prometteur. Les deux idées essentielles que Fabrice a développées sont à la fois l'exploration beaucoup plus profonde des sources d'information dans un moteur thématique que dans un moteur généraliste, mais surtout la valeur ajoutée des moteurs thématiques transformant l'information en objets métiers manipulables. L'exemple qu'il donne d'un moteur touristique illustre très bien ses propos. Imaginez que vous deviez organiser un voyage en utilisant les moteurs généralistes: Vous recherchez un hôtel: plusieurs millions de réponses, vous recherchez un restaurant, encore plusieurs millions de réponses, vous recherchez des sorties, visites, activités et ce sont encore plusieurs millions de réponses. Il faut ensuite croiser toutes ces réponses afin d'organiser votre séjour. Tâche fastidieuse! Un moteur thématique va vous permettre de formuler une demande précise sur vos besoins: "Je veux passer une semaine du 10 au 16 octobre à Chartres en hôtel et réserver un restaurant différent chaque soir, ainsi qu'organiser quelques visites de monuments". Le moteur thématique va alors rechercher les hôtels disposant de chambres libres à Chartres pendant la période demandée. Rechercher pour chacun des soirs de votre séjour les restaurants de la région ayant encore des couverts disponibles et enfin rechercher les visites ouvertes au public à cette saison. Il peut finalement croiser tous les objets ainsi récupérés, les confronter, les organiser, les manipuler et vous fournir en réponses diverses suggestions de séjours... Pas mal non?

Web 2.0

Franck, Sébastien et moi même avons beaucoup insisté sur les mutations en cours. Mutations prenant aujourd'hui la dénomination très controversée de Web 2.0, mais mutations beaucoup plus larges qui vont modifier de manière profonde et significative notre utilisation du Web ainsi que notre poste de travail. Du WebTop de Franck, sorte de boîte renfermant un WebOS jusqu'à mon Wiikle (convergence du Web 2.0, du Wiki, du Web et du Desktop), les deux approchent se rejoignent pour doucement faire pénétrer les technologies Web au cœur même de notre poste de travail ("le navigateur en tant que plate-forme"). Les moteurs devront s'adapter, les moteurs changeront. Ils deviendront ce qu'il sont finalement vraiment: un Service Web, bref, une API. de micro-formats sont certainement ceux qui sont revenus le plus souvent: Ce sont de nouvelles manières de décrire et de structurer l'information, nouvelles manières mettant au centre l'individu et ses réseaux. Au delà des blogs, les moteurs de recherche vont devoir intégrer ces nouvelles méta-informations dans leurs processus d'indexation et de recherche.

Le Point,

Sort un dossier spécial Google le 3 novembre 2005 :

Le monde selon Google

En à peine sept ans, Google est devenu une formidable machine à fabriquer des dollars. Mais le moteur de recherche, qui s'invite également dans l'e-mail ou le téléphone mobile et qui veut même réaliser une bibliothèque universelle, prend des airs de « Big Brother ». Faut-il en avoir peur ?

Dossier réalisé par Etienne Gernelle, François Malye et Armelle Vincent-Arriola (dans la Silicon Valley)

Surtout, ne pas se fier à leurs têtes d'étudiants attardés, de gentils surdoués à peine sortis de leurs jeux vidéo. L'Américain Larry Page et le Russe Sergey Brin, les cofondateurs de Google, font aujourd'hui bien plus peur que les pires requins de Wall Street. A cause de la puissance de leur business, d'abord. En sept ans à peine, le moteur de recherche, lancé dans un garage, a réussi à devenir le numéro un incontestable. Introduit en Bourse il y a à peine plus d'un an, il y pèse déjà plus lourd que General Motors et Ford réunis ! Mais, s'ils font peur, les deux anciens de l'université Stanford doivent s'en prendre à leur prétention à régenter l'information mondiale. Loin du cliché des inoffensifs rêveurs, Page et Brin sont désormais perçus comme des « Big Brothers » en devenir. « *Relax, Bill Gates, c'est au tour de Google d'être le méchant* », titrait récemment le *New York Times*. Google paraît en effet capable de supplanter un jour Microsoft et de lui ravir son statut de grand monopolisateur. Son emprise est déjà impressionnante : disponible en 116 langues, le moteur de recherche est utilisé jusque dans les régions les plus reculées du globe. Impossible pour autant de connaître le nombre de recherches effectuées chaque jour sur son site Internet. Un secret d'Etat chez Google. Le dernier chiffre date de 2003 : à l'époque, 250 millions de questions étaient posées quotidiennement au moteur de recherche, dont la moitié en provenance de l'étranger.



Offre Spéciale

Pour vous,
cette radio-horloge
de voyage...

+ 2 autres cadeaux très utiles

Avec votre abonnement **Le Point**

Le business de l'entreprise se résume à ceci : tout savoir ! Collecter toutes les connaissances mondiales, organiser le fruit de sa pêche interminable grâce à son réseau secret d'ordinateurs - on parle de 175 000 machines - gardés dans un lieu mystérieux, et rendre le tout accessible. Une connaissance universelle qui a une valeur inestimable pour les annonceurs : les liens commerciaux, c'est-à-dire payants, présents sur chaque page mettent directement en contact les entreprises avec les internautes en fonction des recherches de ces derniers. Un système diaboliquement efficace. Pour alimenter son coffre à connaissances, la soif de l'entreprise californienne est sans fin. Après avoir mis en boîte l'ensemble du Web

mondial, Google entend numériser, pour son projet de bibliothèque universelle, Google Print, tous les livres de la planète. Et ce n'est que le début. Pour achever sa conquête de l'univers de l'information, la compagnie a l'intention de maîtriser toutes les technologies de communication : téléphonie, télévision, vidéo, le tout en triomphant enfin du

Peut-on vivre sans Google

11 octobre 2005

30/37

problème de la traduction. Google pourrait même mettre à mal l'hégémonie du système d'exploitation Windows, de Microsoft, en proposant des machines à 100 dollars qui permettraient aux utilisateurs de travailler directement sur Internet. L'ambition de Google paraît n'avoir aucune limite. Les rumeurs les plus folles alimentent déjà sa légende. « *On lui prête par exemple l'intention de se lancer dans la bio-ingénierie* », affirme John Battelle, l'auteur du livre « *The Search* », l'enquête la plus complète à ce jour réalisée sur Google.

La fulgurante ascension de la compagnie commence à inquiéter les défenseurs des libertés individuelles, soucieux du nombre sans précédent de données personnelles amassées et stockées par Google (*voir interview de Kevin Bankston*). Pages Web, mails envoyés ou reçus via Gmail, et même historiques des recherches des individus ! Rien ne dit que toutes ces informations virtuelles définitivement enregistrées n'atterriront pas un jour entre les mains d'agences gouvernementales trop zélées.

Le Patriot Act, voté dans la foulée du 11 septembre 2001, peut ainsi contraindre Google à transmettre les données des utilisateurs au gouvernement. Dans ce cas, la compagnie serait tenue de garder le silence sur cette perquisition informatique. Interrogé à ce sujet par John Battelle, Sergey Brin répond : « *Je n'ai pas lu le Patriot Act, mais je pense que ces inquiétudes sont exagérées. Le gouvernement devrait au moins nous communiquer la nature de sa requête. Je ne pense pas que ce soit un problème sérieux et, si ça le devenait, nous changerions notre politique.* » En attendant, le credo des dirigeants de Google interrogés par *Le Point* reste le même : « *Si nous considérons que ce que nous faisons est bénéfique pour le monde, c'est donc que ça l'est* », assure Richard Chen, l'un des tout premiers dirigeants de l'entreprise. « *Faites-nous confiance. Nous savons ce que nous faisons. Si vous vous y opposez, c'est que vous n'avez rien compris* », ajoute pour sa part Adam Smith, le directeur du projet Google Print, qui a fait décoller la polémique sur Google.

Lorsque les noms de Google ou de ses deux cofondateurs sont prononcés dans la Silicon Valley, trois épithètes reviennent toujours : arrogants, idéalistes et naïfs. « *J'ai travaillé avec eux et je peux vous dire qu'ils sont sincèrement convaincus d'avoir raison et d'avoir transformé le monde*, lance l'ancien PDG d'Ask Jeeves, un moteur de recherche concurrent et partenaire de Google. *Ils affichent un mépris non déguisé pour les opinions contraires aux leurs. Mais s'ils ont une technologie extrêmement sophistiquée, leur vision du monde est très étroite. Ils sont un pur produit du système éducatif américain. Très pointus dans un domaine et ignorants pour le reste.* » Cependant, rares sont ceux qui osent ouvertement persifler sur Google. C'est le meilleur employeur de la région et, vu le rythme de ses embauches, environ 100 par semaine, et le spectre toujours plus étendu de ses activités, chacun peut espérer un jour travailler pour ou avec le moteur de recherche.

Imperméables aux critiques du monde extérieur, Page et Brin tracent leur route. Si l'entrée dans le dictionnaire Oxford est un gage d'immortalité, Google a inscrit son nom dans la postérité. La signification initiale du mot (en réalité, « googol »), qui désigne le nombre 1 suivi de 100 zéros, a été remplacée par une nouvelle définition : « *Utiliser un moteur de recherche, particulièrement google.com* ». De quoi nourrir des rêves mégalomanes. « *Il y a une plaisanterie dans la Silicon Valley qui raconte que le but de Google est de vous faire conduire votre Google vers Google, en vous faisant passer par Google et ainsi de suite*, raconte Gary Rivlin, un journaliste du *New York Times* en poste dans la région. *Tout ce que touchent Larry et Sergey devient de l'or. Et ils n'ont jamais connu d'échecs en suivant leur instinct. Leurs intentions ne sont pas fondamentalement mauvaises, mais ils pensent sincèrement que, s'ils sont écoutés et compris, le monde devien-dra meilleur. C'est très inquiétant.* »

Les multiples polémiques soulevées par l'hégémonie économique, financière et philosophique croissante de Google semblent à peine effleurer les occupants du Googleplex, à Mountain View, près de San Francisco. Un ensemble de bâtiments ultramodernes construits autour d'une place immaculée et verdoyante. Piscine à vagues artificielles, terrain de volley-ball, fauteuils masseurs, piano, toilettes japonaises à jets d'eau tiède chatouilleurs, cabinet médical et navettes (les « googlers » sont écologistes) assurent aux employés un confort qui les maintient dans un contentement béat et les isole d'un monde qui, contradictoirement, semble peu les intéresser. « *C'est typique de la culture de la Silicon Valley, créée autour des ingénieurs informatiques, continue Gary Rivlin. Ils ne lisent pas les journaux. Leur mentalité est la suivante : ce que nous faisons est si important que nous n'avons pas besoin de savoir ce qui se passe. Nous sommes trop occupés à réinventer le monde.* »

Courtois et visiblement passionnés par une mission qu'ils qualifient de purement altruiste, ils refusent poliment de répondre aux questions gênantes et s'en sortent par une pirouette : « *Les utilisateurs peuvent à tout moment passer à un autre moteur de recherche, qui n'est qu'à un clic. Et personne n'oblige personne à utiliser nos services.* » Seul Adam Smith s'énerve lorsqu'on lui demande son avis sur la procédure judiciaire intentée contre Google par l'association The Authors Guild, qui crie au vol des droits d'auteur depuis le lancement des programmes de numérisation Google Print et Google Library : « *Les auteurs ne comprennent pas ce qui est bon pour eux* », répète-t-il. « *J'ai récemment rencontré des gens de Google pour leur soumettre une idée, raconte un entrepreneur de la Silicon Valley qui préfère lui aussi garder l'anonymat. J'en suis sorti avec l'impression que j'avais visité un Etat totalitaire. C'est comme si tous les employés étaient tellement contents de travailler pour Google qu'ils sont devenus débiles. Personne ne veut faire de gaffe.* »

Et personne ne veut non plus être malfaisant, selon le slogan préféré de Google, « *Ne faisons pas le mal* », que Larry Page et Sergey Brin ont d'abord fait afficher sur les murs du Googleplex puis figurer dans le document d'introduction en Bourse. Un principe qui se heurte pourtant aux activités de l'entreprise en Chine. En février 2004, Google pénétrait le marché du géant asiatique en lançant Google News en chinois. Immédiatement banni par le gouvernement, le site n'allait survivre que quelques heures sous sa forme originale avant d'être restauré quelques semaines plus tard, mais cette fois purgé par Google des articles jugés subversifs par le gouvernement chinois. Pourtant, comme l'explique John Battelle, « *la question chinoise pèse sur la conscience de Larry et de Sergey. Après avoir donné des gages de "bonne conduite" à la Chine, ils ont fait venir les meilleurs experts au Googleplex. Selon plusieurs personnes présentes lors de ces réunions, Larry et Sergey avaient une question : comment peut-on rester en Chine sans être complice du mal ? Brin et Page se trouvent désormais dans la position d'adopter le rôle de la police morale de l'économie globale et leurs décisions auront un impact sur les vies de milliards de gens.* »

Le monde selon Google

Picasa : une fois installé sur votre ordinateur, le logiciel Picasa scanne toutes les photos du disque dur et les organise en albums permettant l'envoi, l'impression et la sauvegarde des images.

SMS : toutes les questions peuvent être posées et les réponses reçues en SMS sur votre téléphone portable.

Alerts : il suffit d'inscrire le nom d'un sujet ou d'une personne pour voir tomber dans sa boîte mail toutes les informations le concernant disponibles sur le Web ou dans les 4 500 journaux scannés en permanence par Google.

News : 4 500 journaux et fils d'agence scannés chaque jour. Résultat, une somme considérable

Peut-on vivre sans Google
11 octobre 2005

32/37

d'informations disponible et organisable selon ses propres goûts.

Search : le moteur de recherche le plus utilisé au monde avec 250 millions de mots tapés chaque jour. Des liens commerciaux apparaissent à droite de la page de résultat en fonction des mots demandés.

Mobile : avec la 3e génération de portables, tout Google est accessible d'un téléphone portable. Un marché stratégique pour l'avenir.

Desktop : un logiciel qui sert de moteur de recherche sur votre propre ordinateur. Seule inquiétude : Google a alors accès à tout ce qu'il y a sur votre disque dur.

News : s'affranchir de la frontière de la langue est l'un des objectifs de Google. En partant des mots contenus dans son immense base de données, il devrait révolutionner le résultat des recherches en les traduisant dans la langue de son choix.

Earth : incroyable ! La Terre en images satellites, avec une précision stupéfiante (pas partout, toutefois). Les applications commerciales sont infinies. Une fois visualisée la tour Eiffel, par exemple, le but est de faire apparaître les cafés, hôtels ou boutiques avoisinants sur la page.

Talk : si la messagerie instantanée de Google - le « chat » - n'est pas parmi les leaders de ce segment, elle demeure indispensable à la fidélisation de la clientèle.

Froogle : indispensable à la panoplie de conseils fournis par Google, Froogle est un comparateur de prix.

Print : c'est le projet le plus polémique de Google : numériser l'ensemble des livres de la planète. Un projet pharaonique qui rencontre l'opposition farouche des éditeurs et des auteurs.

Gmail : un service de courrier avec une capacité de stockage stupéfiante : 2,5 Go ! Seul

problème, les mails sont scannés, afin de mieux cerner les goûts de l'utilisateur, et de mieux ajuster les pubs qu'on lui envoie...

Docteur Page et mister Brin

Quand on aperçoit Larry Page en train de discuter avec son ingénieur en chef dans l'un des bureaux de verre du Googleplex, rien dans son maintien, sa tenue décontractée ou ses traits candides ne trahit l'étendue de son pouvoir ni son immense richesse. Cheveux noirs coupés ras et allure de premier de la classe, il est pourtant, à 32 ans, à la tête d'une fortune de 11 milliards de dollars, qui le place au 16e rang des Américains les plus riches. Tout comme Sergey Brin, l'autre fondateur de Google. Qui sont-ils ? Tous deux sont des purs produits du monde académique américain : Carl, le père de Larry, fut l'un des pionniers de l'intelligence artificielle et son épouse enseignait la programmation informatique. Né à Moscou, Sergey est le fils d'un mathématicien et d'une chercheuse à la Nasa, émigrés aux Etats-Unis en 1979. Mais, pour cause de réussite fulgurante, ils n'ont ni l'un ni l'autre terminé leur doctorat à Stanford.

L'argent, Larry et Sergey prétendent s'en moquer. « *Ce n'est pas ce qui les motive* », explique John Battelle, l'auteur du livre « *The Search* », qui retrace l'histoire de Google, et l'un des seuls journalistes à avoir interviewé les deux hommes récemment. « *Larry et Sergey sont des idéalistes.* » Dans un souci écologique de bon ton dans la Silicon Valley, Page et Brin conduisent des voitures électriques lorsqu'ils ne se déplacent pas en patins à roulettes. Ils célèbrent leur succès au Burger King et vivent une existence modeste. On ne leur connaît ni fiancées exubérantes ni goûts extravagants. Personne ne sait trop où habite Larry, mais Sergey n'a pas quitté le deux-pièces qu'il loue depuis ses années d'étudiant à Stanford **A. V.-A.**

Avocat de l'association The Electronic Frontier Foundation

Google représente-t-il un danger ?

Kevin Bankston : Les libertés sont gravement menacées par la capacité de Google d'accumuler un nombre massif de données sur les internautes. Google enregistre et stocke pour un temps indéterminé l'ensemble des activités de ses utilisateurs. Ces données permettent de créer les portraits individuels les plus détaillés jamais conçus. Grâce à l'historique de vos recherches, Google connaît vos habitudes et vos intentions d'achat, votre appartenance religieuse, vos idées politiques, vos amis. Et la compagnie a carte blanche sur la manière dont elle contrôle ces renseignements très personnels.

On prête en général à Google de bonnes intentions. Qu'en pensez-vous ?

Les gens de Google essaient d'appliquer leur principe de base qui est « Ne faisons pas le mal ». Il est difficile d'adhérer à ce credo lorsque vous êtes une entreprise avec des actionnaires à satisfaire. Pour servir toujours mieux les utilisateurs, Google doit collecter de plus en plus de renseignements sur eux. Son modèle commercial est fondé sur l'observation des internautes. C'est dangereux. Depuis l'adoption du Patriot Act, n'importe quelle agence gouvernementale peut s'emparer de ces renseignements dans le plus grand secret.

Quelles sont les solutions ?

Lorsque Google est attaqué sur ces sujets, la compagnie répond : « Faites-nous confiance, nous protégeons vos données. » Mais rien dans ses conditions d'utilisation ou sa politique de confidentialité (1) ne garantit une telle protection. Nous avons besoin de mettre nos lois à jour et les utilisateurs doivent demander des garanties empêchant Google de stocker leurs données. Je conseille aux internautes d'éviter d'utiliser les services de Google comme Gmail, la barre d'outils, Orkut ou le Blogger. Ou bien qu'ils suppriment les cookies et masquent leur adresse IP après chaque session
Propos recueillis par A. V.-A.

Louis Monier : « dommage, en France on ne sait pas prendre de risques. »

Louis Monier, un Français de 49 ans, est reconnu dans la Silicon Valley comme le « père » du moteur de recherche. En 1995, alors qu'il travaille comme chercheur pour Digital Equipment Corporation, il lance un projet ambitieux : la conception et l'écriture du code du premier moteur de recherche, AltaVista. Après un passage par eBay, il vient de rejoindre Google, sans avoir de mission précise. « Je fais le tour des milliers d'idées avant de m'engager sur un projet », dit-il.

Pourquoi la France est-elle en retard ? Louis Monier : Du point de vue téléphonie, la France est plutôt en avance. Mais pour le Web, c'est vrai qu'elle est à la traîne. Historiquement, le Minitel a retardé l'entrée du Web. On avait innové, et puis on s'est arrêté là. Et, pendant ce temps-là, les Etats-Unis nous ont rattrapés, puis devancés.

Manque de matière grise ? Absolument pas. Il y a autant de bonnes idées en France au mètre carré que dans la Silicon Valley. On a des gens extrêmement malins et un système éducatif performant, mais il n'y a pas de terreau. Il est impossible de transformer les idées en quelque chose de concret. Le problème de la France, c'est qu'il est impossible d'y prendre des risques. Les mécanismes en place s'y opposent.

Quels mécanismes ? Les entrepreneurs ont peur de trébucher parce qu'on ne leur donne jamais une seconde chance. S'ils se plantent une fois, adieu tout nouveau financement. Alors qu'aux Etats-Unis un premier et même un deuxième échec sont considérés comme une bonne expérience. Avec de bonnes idées, le financement n'y est jamais un problème. Prenons l'exemple des universités. A Stanford, on met à la disposition des étudiants des laboratoires de recherche et de l'argent. Si une université française proposait de faire ça, on aurait tout de suite une levée de boucliers. Ce n'est pas pour rien qu'il y a entre 30 000 et 40 000 Français dans la Silicon Valley
Propos recueillis par A. V.-A.

Le match des « Big Brothers »

Google veut tout savoir de nous. Microsoft, Yahoo !, eBay... aussi. La bataille a commencé.

Etienne Gernelle (avec Marie-Christine Morosi)

Vous êtes assis à la table d'un café. Avec votre ordinateur portable, vous avez accès à un réseau wi-fi (Internet sans fil) gratuit. C'est gentil, et c'est Google qui vous l'offre. Curieusement, les pubs qui s'affichent à l'écran vantent le restaurant d'en face, le club de gym de la rue d'après. Pas de doute, vous êtes localisé. Pour peu que vous ayez fait des recherches sur Google ou que vous ayez votre mail sur Gmail, tout cela a été scanné, analysé, et permet de cibler mieux encore la pub. Vous aurez droit à celle pour la pizza à croûte fourrée au fromage, votre préférée, et celle pour le jouet « Star Wars » que votre fils vous a réclamé par e-mail cet après-midi. Cela tombe bien, le magasin est sur votre chemin... Un cauchemar ? Non, la réalité déjà, à San Francisco, où Google teste son programme wi-fi.

Offre Spéciale
Pour vous,
cette radio-horloge
de voyage...
+ 2 autres cadeaux très utiles
Avec votre abonnement **Le Point**

Offre Spéciale
Pour vous,
cette radio-horloge
de voyage...
+ 2 autres cadeaux très utiles
Avec votre abonnement **Le Point**

« *Le business de Google, c'est vous* », résume Xavier Dalloz, un consultant spécialisé en technologies de l'information. Un business colossal. Google n'est d'ailleurs pas le seul à en profiter, même si pour l'heure c'est lui qui a le vent en poupe. Microsoft, Yahoo!, AOL, et d'autres encore sont sur les rangs. La guerre des « Big Brothers » a commencé. Tous se battent pour un seul Graal : la connaissance des produits que les gens achètent, des musiques qu'ils écoutent, des amis qu'ils fréquentent. Bref, de leur vie. Les données personnelles valent de l'or.

La fonction de base de Google, la recherche, l'illustre merveilleusement. Si l'on tape la requête « voiture », une publicité pour un constructeur apparaîtra sur la droite de l'écran, dans l'emplacement réservé aux « liens sponsorisés ». A l'aide de l'adresse IP, Google, qui connaît votre lieu de connexion, vous propose une publicité plus ciblée. Exemple : un concessionnaire du Var peut demander que ses liens sponsorisés n'apparaissent que lorsqu'un internaute connecté dans son département tape « voiture ».

Autre avantage, l'annonceur ne paie que si l'internaute clique sur le lien. « *Quand nous démarchons un client, nous savons quel est*

le nombre de recherches réalisées dans son domaine, nous pouvons estimer le nombre de clics, le profil des clients et, in fine, le retour sur investissement, explique Mats Carduner, le patron de Google en France. C'est plus transparent et plus pertinent que tout autre support. Et c'est même plus que du marketing, c'est un outil de vente : à la suite du clic, il peut y avoir un achat en ligne. » L'arme commerciale absolue. Début

Peut-on vivre sans Google
11 octobre 2005

35/37

octobre, Eric Schmidt, le PDG de Google, l'a décrite avec minutie devant le congrès annuel des annonceurs américains, sous l'oeil sidéré des représentants de médias traditionnels.

Le système Google comporte toutefois une exigence : il faut contrôler tous les canaux, encercler le « client ». L'e-mail, les messageries instantanées, mais surtout ce qui est mobile. « *Nous sommes à l'ère de l'Atawad*, souligne Xavier Dalloz : "*Anytime Anywhere, Any Device*". En français : *n'importe quand, n'importe où, et quel que soit l'outil*. » Certaines options ne rapportent que peu d'argent, mais elles servent à conserver le client à portée de pub. « *Le but est qu'on ne se détache pas de Google*, explique Franck Poisson, ancien patron de Google France, parti chez un moteur concurrent, Exalead.com. *Par exemple, si le client a un mail chez Yahoo, il risque d'y faire aussi ses recherches Internet*. » Moralité, pour ne pas courir le risque d'être court-circuité par un concurrent, il faut l'étouffer d'abord. Si Google a lancé si tôt son incroyable Google Earth, alors qu'il n'en est visiblement qu'au stade de test, c'est parce que Bill Gates a imprudemment évoqué un projet similaire dans une interview...

La course est impitoyable, mais, pour l'heure, les inventions des apprentis Big Brothers ne fonctionnent pas toutes à plein régime. « *Il y a deux fantasmes sur Google*, analyse Franck Poisson. *L'un est vrai, l'autre est faux. Celui qui est vrai, c'est Google Big Brother. Oui, ils ont votre adresse IP ; oui, ils connaissent votre historique de recherche et lisent vos mails. Tout cela est vrai. En revanche, ils ne peuvent pas tout faire en même temps. Par exemple, leur alliance avec Sun pour promouvoir un système d'exploitation concurrent de Windows, je n'y crois pas beaucoup*. » Il reste donc un répit.

L'autre souci, pour les Big Brothers en herbe, c'est qu'ils finissent toujours par agacer. Ainsi, la volonté hégémonique de Microsoft lui vaut depuis des années une réputation exécrationnelle. La firme de Bill Gates a réalisé il y a six mois une enquête pour comprendre pourquoi tant de gens honnêtes pirataient leurs logiciels. Réponse : parce qu'ils se sentent contraints. Pour Google, le risque de se « microsoftiser », comme on dit sur les blogs dédiés à la question, n'est pas négligeable. Il est toutefois peu probable que l'entreprise de Mountain View abandonne sa stratégie. « *Le pouvoir absolu que Google s'est donné corrompt*, estime Philippe Coueignoux, président de eprio, une start-up spécialisée dans la confidentialité sur Internet. *Le désir de croissance les pousse à aller de plus en plus loin*. »

Le duel entre Microsoft et Google risque de faire du bruit dans les mois et années à venir. Moins fringant pour le moment, le géant de Redmond a des ressources. Donné pour battu à plusieurs reprises, notamment par Netscape en 1995, Microsoft a toujours fini par écraser l'adversaire. Le géant de Redmond pourrait sortir l'année prochaine un moteur de recherche concurrent. « *La bataille n'est pas gagnée pour Google*, assure Franck Poisson. *Ce n'est pas la fin de l'histoire*. » Quoi qu'il en soit, tout cela reste une bataille entre l'Etat de Washington et la Californie. L'Amérique domine le monde de l'information comme jamais. Et la seule entreprise européenne du secteur qui avait réussi de manière insolente, Skype (1), spécialiste de la téléphonie sur Internet, vient d'être rachetée par l'américain eBay. « *L'affaire Skype est le drame de l'Europe* », affirme André-Yves Portnoff, un chercheur spécialisé dans les nouvelles technologies, qui pointe l'absence de réel capital-risque, des mentalités conservatrices, mais aussi un problème légal. « *Les Etats-Unis ont le Small Business Act, rappelle-t-il, en vertu duquel une partie des marchés publics est réservée aux PME. Cela aide les start-up à grandir. Chez nous, soit elles restent naines, soit elles sont rachetées et étouffées*. » L'Europe étant aux abonnés absents, seule l'Asie semble en mesure de contrer l'Amérique. « *Par son marché, la Chine va pouvoir imposer ses standards* », prédit André-Yves Portnoff. Google, Microsoft et Yahoo ! se livrent une bagarre féroce en Chine, n'hésitant pas à débaucher les ingénieurs des autres. Le pouvoir se déplace à l'Est. De là à imaginer un jour un Big Brother multipolaire...

Peut-on vivre sans Google

11 octobre 2005

36/37

Google en chiffres

Effectifs : 5 000 salariés.

Chiffre d'affaires 2004 : 3,1 milliards de dollars. Résultat net 2004 : 399 millions de dollars.

Chiffre d'affaires du 3e trimestre 2005 : 1,58 milliard de dollars, en hausse de 96 % par rapport à l'année précédente.

Cours de l'action : 357 dollars au 28 octobre 2005, contre 100 dollars à l'introduction, en août 200

A la conquête des petites annonces

C'est la dernière trouvaille de Google : un service - encore au stade de test - baptisé « Google Base », qui permettrait aux internautes de rendre accessibles en ligne leurs petites annonces, que ce soit des faire-part ou pour vendre des objets d'occasion. « *Nous sommes en train d'expérimenter de nouveaux moyens pour que les propriétaires de contenus puissent les envoyer facilement à Google* », précise le groupe. Google Base pourrait toucher les revenus de la presse traditionnelle, mais aussi être une menace pour eBay, le site d'enchères en ligne qui, contrairement à Google base, demande une cotisation à ses vendeurs. La gratuité est encore une fois l'arme absolue de Google. Cette nouvelle fonctionnalité devrait permettre à Google d'augmenter encore le trafic sur son site, mais aussi de fournir de précieuses données sur les vendeurs et les acheteurs. De l'or en barre pour les annonceurs **E. G.**

1. Fondée par un Danois et un Suédois, Skype avait son siège au Luxembourg.

© le point 03/11/05 - N°1729 - Page 85 - 1100 mots