

Rapport publié le 06 septembre 2007

23 octobre 2006

**POUR UNE INITIATIVE EUROPÉENNE DANS L'INTERNET EN FAVEUR DU
DÉVELOPPEMENT DE L'ÉCONOMIE DE LA CONNAISSANCE**

QUE FAIRE FACE À GOOGLE ?

Plus qu'au contrôle des outils de traitement et de gestion de l'information (Microsoft, Intel, Cisco), c'est à la maîtrise de l'accès à l'information et de sa diffusion que la France et l'Europe sont confrontées aujourd'hui. L'enjeu ne se limite plus aux investissements dans les réseaux et les supports (ADSL, fibre optique, etc.), mais s'étend aux contenus numériques et impose d'être acteurs dans la formation de ces derniers.

L'ensemble des moteurs de recherche et d'archivage de ce qui transite sur le Web peut-il être sous la responsabilité d'un seul État, fût-il ami ? Est-il économiquement et éthiquement acceptable de dépendre, pour une fonction aussi importante que l'accès à l'information, quasiment d'une seule société ? Comment encadrer par les règles habituelles du marché des prestations mises gratuitement à la disposition de centaines de millions d'utilisateurs ?

L'émergence d'un nouveau modèle économique d'opérateurs, prêts à offrir aux internautes tous les services d'accès gratuitement à condition que ceux-ci mettent à leur disposition leurs informations personnelles, place les décideurs nationaux et européens face à des défis juridiques et politiques majeurs.

1. UNE DÉPENDANCE GRANDISSANTE VIS À VIS D'UN PETIT NOMBRE D'OPÉRATEURS

Google est l'archétype de ce nouveau modèle d'opérateur. Non seulement Google représente le premier moteur de recherche et de loin le plus performant, mais il constitue une entreprise hors normes, d'une puissance économique considérable. Google pèse aujourd'hui environ 100 milliards d'euros, et son modèle d'affaires, fondé sur une offre a priori gratuite pour l'utilisateur, constitue une approche qui ne peut que déstabiliser l'ensemble des acteurs en place.

1.1. LES SERVICES DE GOOGLE

Son image de spécialiste de la recherche sur Internet est trompeuse : Google ne se contente pas, loin de là, d'un positionnement apparemment inoffensif sur ce qui pourrait apparaître comme une niche de services d'Internet. En louant les espaces publicitaires¹ que son moteur de recherche génère, Google dispose d'une rente de situation qui lui donne les moyens d'être présent sur l'ensemble des services Web. Le modèle d'affaires est reproduit dans ces nouveaux services, qui emportent une petite dose de publicité, mais sont de plus en plus éloignés du moteur de recherche qui a fondé la puissance économique de Google.

En effet, dès aujourd'hui, en plus de sa fonction de moteur de recherche, Google offre gracieusement les fonctions décrites en annexe², pour la plupart très performantes.

[Google Base](#) propose aux particuliers et aux entreprises d'y déposer toutes les données qu'ils souhaitent rendre publiques, quelles que soient leur structure ou leur complexité : petites annonces en tous genres ou catalogue complet des produits d'une entreprise ...

¹ Le marché publicitaire en France a été en 2004 de l'ordre de 31 Milliards d'euros soit pour chacun des 26 millions de ménages une dépense de 1.200 € annuellement! Si seulement 5% de cette somme passe aujourd'hui par Internet, on comprend que réinvestir une partie de cette somme dans le média Internet va devenir rapidement rentable.

² Version anglaise car plus complète

C'est ainsi la possibilité de mise en œuvre et de contrôle de véritables canaux de vente et d'échanges qui est visée. C'est aussi déjà Google Checkout : c'est-à-dire un système de micro-paiements qui risque de devenir une norme de fait, en l'absence d'entente et de proposition du système bancaire traditionnel.

De surcroît, Google Drive, c'est-à-dire la mise à disposition d'un immense disque dur virtuel, est annoncé pour bientôt. Au départ, il s'agira d'offrir à chacun une sauvegarde des données stockées sur son PC. Mais, de fait, il apparaîtra très rapidement que ce disque sera le disque de référence des clients. Il sera en effet extensible à l'infini et accessible à partir de n'importe quel poste (PC, téléphone, etc.).

1.2. LES ATOUTS TECHNOLOGIQUES DE GOOGLE

Pour mettre en œuvre son outil de recherche, Google a fait un choix technique très significatif : équiper ses serveurs de processeurs, issus du commerce, et les empiler en se donnant comme unique contrainte de pouvoir redistribuer la puissance de calcul et de stockage sur ses différentes machines en fonction des besoins, en rajoutant « à la volée » autant de machines que le trafic dans une région donnée l'exige. Les experts les évaluent à quelques 450.000 serveurs, répartis dans une vingtaine de « data centers ». Ce faisant Google a construit une architecture informatique propriétaire sans équivalent, très souple et très robuste et une phénoménale capacité de calcul qui permet de répondre à plus d'un milliard de requêtes par jour, chacune interrogeant huit milliards de pages Web en moins d'un cinquième de seconde...

Il s'agit à l'évidence d'un atout important de Google. La puissance informatique installée permet une adaptation à la variation de la demande, le développement de nouveaux services et une qualité de service de bon niveau. Les équipes techniques de Google ont su maîtriser une architecture d'informatique distribuée mâtinée de calculateurs centralisés (les « data centers »). Elles ont développé un savoir-faire allant de la programmation des instructions élémentaires des micro-processeurs à l'optimisation des fonctions logicielles de management des systèmes en passant par la maîtrise d'une architecture évolutive et standardisée, qui conduisent à une articulation particulièrement efficace entre algorithmique et architecture. Ce savoir-faire pourrait bien constituer la véritable avance technologique de Google.

Les coûts correspondants sont remarquablement bas (450.000 machines soit environ 200 M€). Google peut ainsi loger sur ses serveurs non seulement les index et les copies du Web, mais aussi tout le reste : c'est-à-dire l'intégralité des données importantes (10 % des disques durs) des ordinateurs de ses utilisateurs !

1.3. LE POTENTIEL DE GOOGLE

Ainsi :

- Google a mis en œuvre une architecture et un système d'exploitation qui lui permettent à moindre coût de « tout » stocker ;
- Il offre gratuitement cet espace à tous les utilisateurs ;
- Il offre également les outils pour consulter mais aussi exploiter ces données de manière simple et partageable dans la sphère privée ou professionnelle ;
- Google finance sa démarche avec les profits de la publicité pour laquelle il développe de nouveaux marchés et attire de nouveaux annonceurs (cf. la théorie de la « longue queue »³) ;
- Son modèle économique basé sur la publicité et la gratuité l'expose moins que Microsoft au droit de la concurrence ;
- Il est important de noter que le fait que ces fonctionnalités soient mises gratuitement à la disposition des utilisateurs s'accompagne d'une exigence de qualité de service de bon niveau, qu'il s'agisse de l'ergonomie des écrans, des temps de réponse ou de l'étendue des fonctions annexes et des liens offerts. Google met ainsi largement en œuvre le principe qui voudrait que, dans une économie de la connaissance, l'accès à l'information soit gratuit ;

³ « LongTail » fait référence à la forme d'une courbe commentée par le [management de Google](#) montrant qu'en touchant des millions de niches, Internet peut automatiquement satisfaire des micro-marchés dont la somme est un énorme marché nouveau

- Il accélère son emprise économique avec un rythme exceptionnel de sortie de produits quasi hebdomadaire. Certes, avec une granularité des services qui reste fine, les fonctionnalités offertes sont largement indépendantes les unes des autres. La perspective d'une intégration et d'une mise en cohérence de ces services reste lointaine, ce qui facilite cette profusion de nouveautés et rend moins sensible la perception des risques intrusifs et invasifs dans la vie personnelle des internautes. Cette stratégie est à l'œuvre et il est fascinant de voir qu'elle peut avancer aussi vite, sans aucune contrainte. Or les enjeux, notamment de souveraineté, sont immenses !

2. DES CAPACITÉS EUROPÉENNES MAIS DES RÉPONSES PONCTUELLES ET NON COORDONNÉES

2.1. DES COMPÉTENCES RECONNUES

La France dispose de compétences reconnues sur le plan international, qu'il s'agisse du domaine des principes et algorithmes de recherche d'informations par le contenu, ou de la construction de très grandes infrastructures informatiques capables de fournir les performances d'un site comme celui de Google.

2.1.1. RECHERCHE PAR LE CONTENU

Déjà au cours des années 70 et jusqu'à la fin des années 80, le logiciel Mistral de la société CII/Bull était un leader mondial en matière de recherche documentaire.

L'école française de Recherche se distingue en majorité de l'école américaine par un fort recours à la linguistique pour améliorer la pertinence des réponses aux requêtes. Des équipes structurées existent dans les grands organismes comme le CEA, le CNRS ou l'INRIA, localisées principalement à Grenoble, à Nancy, en région parisienne et à Toulouse. Au-delà du problème de la recherche d'informations textuelles, les laboratoires s'orientent vers la recherche d'informations multimédia (image, son, vidéo). Des équipes du GET ou de l'école des mines de Paris travaillent ainsi sur les aspects de l'image, de la vidéo voire de la reconnaissance biométrique.

A côté de ces équipes de recherche plusieurs « jeunes pousses » se sont créées, qui conservent dans leurs produits cette caractéristique d'une approche linguistique, comme Lingway, New Phoenix, GO-Albert, Sinequa, La société Exalead quant à elle se positionne directement face à Google avec des techniques de moteur de recherche similaires.

En matière de recherche d'images les sociétés LTU ou encore New-Phoenix ont développé des offres. Certains grands groupes ont constitué de fortes équipes sur le sujet, tels Thalès, France Télécom, ou Thomson.

2.1.2. INFRASTRUCTURE INFORMATIQUE

Le rôle de l'infrastructure dans la fourniture de services du type de ceux de Google est essentiel. L'architecture à développer est celle dite d'un cluster qui permet de regrouper de manière cohérente au sein d'une même infrastructure plusieurs dizaines, centaines, ou milliers de serveurs interconnectés avec ce qu'il convient de modularité de croissance (scalabilité), de haute disponibilité, de sécurité et de « manageabilité ».

Deux types de traitements sont mis en œuvre :

- des traitements de type « portail » où il s'agit d'écouler un maximum de requêtes indépendantes en même temps, chaque requête étant simple ;
- des traitements ou requêtes dont la complexité nécessite de paralléliser les algorithmes correspondants pour obtenir un temps de réponse raisonnable.

Pour les traitements de type « portail », les solutions relèvent des constructeurs informatiques ou de quelques grandes sociétés de services qui disposent des compétences techniques nécessaires. Il faut aussi mentionner l'existence de jeunes pousses très innovantes comme Kewego ou Dailymotion dans le domaine en émergence des portails de blogs vidéo.

Pour les traitements parallèles, l'infrastructure est plus complexe et très peu d'acteurs ont les compétences techniques pour aborder les architectures de très grande taille. Exalead a démontré ses compétences « système » sur des configurations moyennes et Bull sur de très grandes configurations dans le domaine du calcul à haute performance. Dans les deux cas des logiciels Open Source peuvent être utilisés, même si l'importance des configurations à mettre en œuvre nécessite une très grande expertise technique pour optimiser ces logiciels et les rendre performants à l'échelle requise.

2.2. DES INITIATIVES LIMITÉES ET NON COORDONNÉES

Il existe en France et en Europe un foisonnement d'initiatives privées relevant des champs décrits en point 1.

On peut citer à titre d'exemples :

- les banques d'images Kewego (<http://www.kewego.fr/>) et Wat (<http://www.wat.tv/>) ;
- la téléphonie sur Internet Wengo (<http://www.wengo.fr/>) ;
- les données géographiques, plans et itinéraires Mappy (<http://www.mappy.com/>) ;
- la traduction automatique avec Systran (<http://www.systran.fr/>).

Cependant, hormis ces deux dernières qui ont acquis une stature internationale, elles restent quelque peu confidentielles et sans lien entre elles. Leur diffusion et leur croissance sont limitées par le manque de capacités matérielles et financières leur permettant d'asseoir leur développement en Europe et par l'absence de grands vecteurs fédérant leur promotion.

Les quelques initiatives publiques européennes et françaises :

- la bibliothèque numérique européenne (<http://www.theeuropeanlibrary.org/>) ;
- le système d'information géographique Geoportail : <http://www.geoportail.fr/> ;
- les archives audiovisuels de INA (<http://www.ina.fr/>) ;
- le moteur de recherche Quaero : <http://www.quaero.fr/> ;

souffrent des mêmes maux et s'ignorent entre elles. Il suffit de se connecter aux sites de la Bibliothèque numérique européenne ou de Quaero pour constater combien les résultats souffrent de la limitation des moyens.

Dans le même esprit, la récente décision (du 29 septembre 2006) du Ministère de l'économie, des finances et de l'industrie de confier à Thalès la mise en place pour fin 2007 d'un service de consultation du plan cadastral pourrait être analysée à l'aune de la même carence de vision d'ensemble.

3. QUE FAIRE ?

3.1. LES MOTIVATIONS POUR AGIR

Les questions de souveraineté sont importantes, même si leur perception par le grand public est variable. En effet, si les États apparaissent légitimes pour le contrôle des informations de nature régaliennne (comme l'état civil par exemple), la situation est toute autre pour les données non publiques. De fait, Google a démontré qu'il avait une légitimité implicite à contrôler des informations de nature privée : d'aucuns préféreraient mettre leur agenda personnel sur Google que de le placer sur un site du ministère de l'Intérieur.