

## Chapitre 15 – Les moteurs de recherche

Les moteurs de recherche ont représenté un service critique du Web dans la mesure où ils fournissent l'interface principale du Web. On peut même dire qu'ils ont joué (et jouent encore pour les vagues de nouveaux utilisateurs) le rôle de "point de départ" dans l'utilisation du Web et de l'Internet pour les utilisateurs non-techniciens (l'immense majorité désormais). Grâce à ce type de service, l'internaute peut en effet trouver des sites Web qui l'intéressent sans avoir à connaître au préalable leur adresse (URL) voire même leur existence !

Les moteurs de recherches sont devenus une pièce critique dans l'économie du Web — surtout lorsque l'on cherche à promouvoir son site. Car pour tous les sites Web de petite ou moyenne taille, être référencé par les moteurs de recherche est un passage obligé pour se faire connaître. D'où une pléthore de termes techniques liés au sujet : SEO (Search Engine Optimization : comment optimiser son site Web pour être en tête des résultats retournés), SEM (Search Engine Marketing : comment être visible sur les moteurs de recherche, que ce soit par le biais du SEO ou en achetant des encarts publicitaires), CTR (Click-Through Rate : pourcentage d'internautes qui cliquent sur l'encart publicitaire), etc.

Le principe du Web étant basé sur les hyperliens entre différentes pages, la notion de moteur de recherche est venue naturellement. Un site Web peut en effet facilement faire office de moteur de recherche pour d'autres sites Web, alors qu'un site FTP ne peut pas faire facilement office de moteur de recherche pour d'autres sites FTP. Le Web a ainsi vite remplacé Gopher, un service de recherche de document sur Internet populaire avant la montée du Web. Le fait que Gopher était moins souple que les moteurs de recherche Web (il exigeait que les documents suivent un format donné), moins facile d'utilisation qu'un navigateur Web et que certains serveurs Gophers ont commencé à être payants a fait que le service est tombé en désuétude rapidement après avoir été à la mode -brièvement- dans les années quatre-vingt-dix. Comme souvent dans des nouveaux marchés, on a tout d'abord assisté à une explosion des moteurs de recherche Web, avant de voir un tassement, les moteurs les plus populaires triomphant au dépend des plus petits.

### Les débuts des moteurs de recherche sur le Web : Yahoo contre AltaVista

Les premiers moteurs de recherche sont apparus très vite. Dès 1993 on a eu Aliweb, W3Catalog ou JumpStation alors que le Web était encore très "jeune". En 1994 ont suivi de nombreux autres moteurs, certains restants encore connus aujourd'hui : Yahoo (compagnie fondée en 1995), Architext (renommé Excite), Lycos, WebCrawler, etc. Ces compagnies furent créées pour de nombreuses raisons et pas forcément pour indexer le Web. Excite a commencé en 1994 par six étudiants de Stanford sous le nom d'Architext et avait pour but de rechercher dans les bases de données des entreprises. Vinod Koshla (cofondateur de Sun), capital-risqueur de Kleiner Perkins Caufield & Byers, était derrière Excite et c'est lui qui suggéra aux fondateurs d'indexer plutôt le Web. Très vite deux moteurs de recherche se sont imposés : Yahoo d'un côté et AltaVista de l'autre.

Cow-boys contre chemin de fer ou que savez-vous vraiment de l'histoire de l'informatique ?

Yahoo a commencé en février 1994 par deux étudiants de Stanford (encore), David Filo et Jerry Yang. Selon la version officielle, le site (alors appelé "David & Jerry's Guide to the World Wide Web") a commencé comme un moyen de répertorier les sites Web qui les intéressaient. Lorsque le site a pris de l'ampleur, ils ont dû créer des catégories. Lorsque les catégories devinrent trop grosses, elles durent être divisées en sous-catégories et ainsi de suite. Le concept de Yahoo est né comme cela. Mais un autre élément de Yahoo qui n'est pas relaté par la version officielle est que le moteur a réellement commencé avant même que Filo et Wang n'utilisent le Web. Dans son livre *The Search*, John Battelle a interviewé Jerry Yang et David Filo qui ont reconnu, à l'époque, être prêts à tout pour ne pas travailler sur leur thèse (la spécialité de Yang et Filo n'était pas l'informatique mais le génie électrique). L'ancêtre de Yahoo fut créé avant que le Web ne se répande, et était un programme qui indexait automatiquement des informations de sites de basket-ball et ce afin de gagner des matchs de fantasy baseball (ce jeu, très populaire aux Etats-Unis, permet aux joueurs de gérer une équipe de baseball virtuelle basé sur les performances de véritables joueurs).

En 1995, les deux étudiants décidèrent de renommer leur site en Yahoo pour Yet Another Hierarchic Officious Oracle. En fait ils décidèrent tout d'abord que le nom devait commencer par YA, suivant une tradition en informatique d'utiliser des noms en "Yet Another". En parcourant le dictionnaire à la lettre Y, le terme Yahoo résonna immédiatement (en anglais, Yahoo veut à la fois dire "Youpi" et "Barjot"). Pour le reste, la page de garde de Yahoo, avec son logo rouge et ses catégories principales est restée célèbre pendant des années.

De l'autre côté, AltaVista, un service lancé par Digital Equipment Corporation (DEC), a pris une approche diamétralement opposée. Yahoo référençait manuellement les sites Web, alors qu'AltaVista effectuait une indexation automatique. Autant le référencement du Web était une fin en soi pour les fondateurs de Yahoo, autant c'était un moyen pour Digital. AltaVista fut créé comme vitrine technologique de la compagnie, le service utilisant des puissants serveurs à base de l'Alpha, le tout nouveau processeur 64-bit de la compagnie. L'origine de l'idée est contestée, mais ce fut Louis Monier, un français travaillant pour Digital au laboratoire de Palo Alto, qui coda la première version. Ce fut également Monier qui réussira à convaincre la direction de DEC de dévoiler AltaVista au grand public (pour la direction, le seul but était de vendre des serveurs) en leur promettant une publicité fabuleuse pour la compagnie. Et, effectivement, Alta Vista fut rapidement très populaire mais il est difficile d'estimer l'impact réel pour DEC car, comme d'habitude avec tout ce qui est radicalement nouveau et hors de sa culture d'origine, DEC a été trop timide dans l'exploitation marketing de la popularité soudaine d'AltaVista.

AltaVista a tout de suite eu une culture très technique et peut être considéré sur de nombreux points comme le précurseur de Google. Tout comme Google, son premier challenge a été d'indexer le Web de manière efficace là où les moteurs de recherche de l'époque avaient de nombreuses carences : mauvaise interface graphique, langage de requête peu puissant et/ou pas assez de pages indexées. AltaVista s'est appuyé sur une puissante architecture matérielle (une douzaine de serveurs DEC Alpha) pour fournir un service qui était bien plus complet que ce qui existait déjà. Là où Yahoo se basait sur un tri effectué par des humains, AltaVista a proposé une recherche par mot-clé. Enfin, AltaVista a dès le début proposé une interface minimaliste -ce que Google a réinventé à ses débuts. En parallèle, de nombreux autres services ont tenté de s'accaparer leur part du gâteau : Excite, Lycos, Ask Jeeves (renommé depuis Ask.com). La plupart se basant sur des mots-clés. Certains, comme Ask Jeeves, se sont basés sur le langage naturel avec plus ou moins de succès. Mais aucun de ces services n'a réellement réussi à écorner le succès des deux géants du moment.

## **Du moteur de recherche au portail (1996-2000)**

Avec la bulle Internet un nouveau mot à la mode est né : le portail. À l'époque de la soi-disant "nouvelle économie", le mot d'ordre était d'attirer le plus de trafic possible sur son site Web. Chercher à monétiser ledit trafic n'était pas à l'ordre du jour. L'opinion communément admise de l'époque était que cela viendrait naturellement, même si à l'époque personne ne voyait comment. Les principales valeurs de l'époque ont donc tourné autour des

mesures qui évaluaient le trafic : nombre de hits, nombre de visites, de visiteurs uniques par mois, etc. A l'époque, même les publicités étaient payées en terme de hits, c'est-à-dire de nombre de fois où elles étaient visualisées (de \$5 à \$20 pour 1000 affichages suivant que le public est ciblé ou non). Ce modèle était calé sur le modèle publicitaire traditionnel où la seconde de publicité télévisée est par exemple facturée proportionnellement à l'audience. Mais il suivait également la mentalité de l'époque : on recherche la quantité de visiteurs plus que la qualité.

Dans ces conditions, la notion de portail Web est rapidement devenue le Graal de la bulle Internet. Tim Koogle, PDG de Yahoo de 1995 à 2001, a défini le portail comme étant le point de passage obligé du Web. La course au portail a tout naturellement attiré beaucoup de monde. De nombreuses compagnies ont tenté de diriger leurs clients ou utilisateurs vers leur propre portail dans l'espoir de devenir le portail le plus important. AOL, le plus gros fournisseur d'accès Internet de son temps, s'est appuyé sur ses clients. Netscape, fort de son navigateur à succès, a fait de même. Yahoo s'est rapidement "réinventé" en portail Web.

Avec la course au portail on a assisté au déclin de la qualité de nombreux moteurs de recherche. Yahoo a été sur ce point un exemple caractéristique, mais AltaVista est également tombé dans ce piège. Car avec le concept du portail, le moteur de recherche n'occupe plus qu'une petite fraction du service. En effet, l'idée est de limiter au maximum le besoin de surfer le Web en proposant au sein d'une seule et même page le plus d'information possible. Les portails génériques (ainsi que de nombreux portails spécialisés) ont donc commencé à afficher des informations telles que l'actualité, la météo ou le cours de la bourse. Les portails ont en plus cherché à intégrer les services les plus utilisés par les internautes. Yahoo a ainsi développé de nombreux services Web. Yahoo Mail bien sûr, mais Yahoo Instant Messaging (messagerie instantanée), Yahoo Personals (site de rencontres en ligne), Yahoo Maps, Yahoo Movies, Yahoo Finance et de nombreux autres services.

Le concept du portail était en fait une évolution naturelle pour une compagnie comme Yahoo. Car dans les deux cas l'idée est qu'une poignée de gens décident quels types d'informations devraient être sur la première page que l'internaute voit en lançant son navigateur Web. A ses débuts, Yahoo affichait sur sa page de garde les catégories de sites Web qu'il jugeait principales. En devenant un portail Web, la compagnie affiche les éléments du Web qu'elle juge primordiales. Mais c'était sans compter l'explosion du Web et de ses utilisations. Le fait que quelques dirigeants décident d'un type de page unique que tous les internautes devraient voir n'a pas tenu longtemps. Essayer de garder l'utilisateur le plus longtemps sur son site Web est contraire à l'esprit du Web -les millions de sites Web indépendants existent justement pour cette raison. Le phénomène de fidélisation à une marque n'a pas fonctionné comme désiré. Du fait de la nature "éclatée" du Web, les internautes n'ont pas adopté des services simplement pour leur marque comme le prévoient et l'espéraient les promoteurs des portails. Google Maps par exemple, a supplanté Mapquest et Yahoo Maps non pas parce qu'il était estampillé Google mais parce qu'il apportait vraiment quelque chose de plus.

Pire, le portail a créé un encombrement des pages de recherche Web. Très peu de gens sont intéressés par tous les éléments affichés. Les éléments qui n'intéressent pas un internaute ne font qu'augmenter le fouillis de la page. L'éclatement de la bulle a fait empirer les choses : les sites Web ayant plus de pression pour être rentables ont donc fait appel à de la publicité souvent agressive. Si AltaVista a depuis abandonné le concept de portail et est revenu à un look minimaliste concentré sur le moteur de recherche (eh oui, AltaVista existe encore : après être passé par de nombreuses vicissitudes, le service fait désormais partie d'Overture), les pages de garde de Yahoo ou AOL sont restées tellement encombrées qu'elles en sont inconfortables. Certains services de Yahoo (comme Yahoo Movies) possèdent deux barres de recherche : une pour chercher les films, et une pour chercher le Web, amenant beaucoup d'utilisateurs à confondre les deux.

=== La difficulté de la diversification

De nombreuses compagnies ont été critiquées pour s'être trop diversifiées et ainsi "s'être éloignée