

Stephen Taylor (Yahoo! Search) : "Nous voulons intégrer l'internaute dans le processus de recherche"

[Retour au sommaire de la lettre](#)

Nous avons rencontré, à l'occasion de l'inauguration du nouveau centre de R&D de Yahoo! à Barcelone début octobre, Stephen Taylor, l'un des responsables européens du moteur de recherche de la firme américaine en Europe. Une bonne occasion pour faire le point avec lui sur les travaux en cours et surtout sur la stratégie globale de Yahoo! en termes de "search" dans le monde.

Yahoo! a inauguré à Barcelone le lundi 9 octobre 2006 son premier centre de recherche et développement en Europe, placé sous la direction de Ricardo Baeza Yates (<http://yhoo.client.shareholder.com/press/ReleaseDetail.cfm?ReleaseID=185028>). A cette occasion, nous avons pu rencontrer bon nombre de responsables du moteur de recherche de la société américaine. Nous avons choisi de porter à votre connaissance les réponses que Stephen Taylor, "Regional vice president and managing director Search and Search marketing" de Yahoo! Europe, a bien voulu faire à nos (nombreuses) questions...

Stephen Taylor a travaillé chez Overture à partir d'avril 2003 en tant que Directeur commercial puis chez Overture Europe en 2005 avant le rachat par Yahoo! de la société qui allait devenir "Yahoo! Search Marketing" par la suite. Il est aujourd'hui responsable pour Yahoo! Europe de l'innovation en termes de contenu et de produits pour la recherche et le "search marketing". Merci à lui pour le temps qu'il a bien voulu nous consacrer.

Et tous nos voeux de réussite pour l'équipe de football d'Arsenal cette année, Stephen (private joke ;-)...



Stephen, bonjour et merci de répondre à nos questions. Comment pourriez-vous définir la stratégie de Yahoo! en termes de recherche d'information sur le Web ?

Notre but tient en une phrase : "*enrich people's lives by enabling them to find, use, share and expand all human knowledge*". En français : "enrichir la vie des internautes en leur permettant de trouver, utiliser, partager et accroître la connaissance humaine." Ce credo est important :

- "Trouver" car nous devons être à même de permettre aux internautes d'identifier la bonne information au regard de leurs besoins.
- "Utiliser" car nous devons proposer des outils pour traiter l'information dans le but d'accomplir une tâche ou réaliser un but.
- "Partager" car nous sommes persuadés que c'est dans l'échange que l'on avance le plus vite. Cette notion est à la base de notre réflexion actuelle.
- "Augmenter" car il est important de fournir aux gens le plus large éventail de réponses et de solutions possible.

Ce concept de "FUSE" (*Find, Use, Share and Expand*) nourrit nos réflexions quotidiennes afin de fournir les meilleurs résultats de recherche possibles à nos visiteurs.

Yahoo! est un réel concurrent de Google aux Etats-Unis et en Asie pour ce qui est du moteur de recherche, mais moins en Europe. Pourquoi selon vous ?

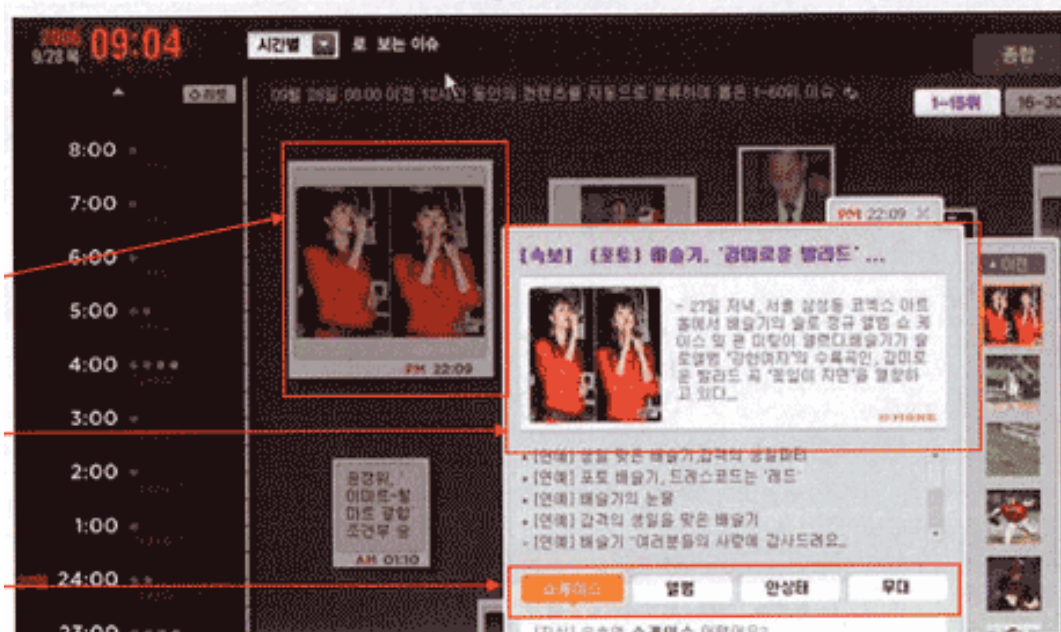
Difficile à dire. Google est certainement arrivé au bon moment, lorsque les places à prendre étaient peut-être plus "faciles" à acquérir. La situation a évolué mais nous restons persuadés que le jeu est loin d'être terminé en Europe, bien heureusement. Pour l'instant, nous essayons de nous focaliser sur ce que nous voulons proposer à nos visiteurs sur ce continent. Nous verrons comment cela évoluera dans les mois qui viennent... Nous restons très confiants.

Il n'est donc pas trop tard ?

Bien sûr que non. regardez la situation en Corée du Sud. Yahoo! y est leader de la recherche d'information, loin devant Google. Nous avons beaucoup innové sur notre portail coréen, avec une notion de "dynamic clustering results", ou présentation des résultats en "nuages d'images" (plus l'image de l'utilisateur est grande, plus le mot clé qui la représente est populaire), avec les possibilités de redéfinir les résultats de recherche par sources (blogs, web, etc.) et de recherche par tranches horaires, etc. Bref, de nombreuses nouveautés appliquées au concept de communauté qui nous ont permis, dans ce pays, de rapidement devenir leader.



La notion de "Dynamic Clustering Results" sur Yahoo! Corée.



La possibilité de recherche par tranche horaire sur ce même site.

La situation est identique à Taiwan où Yahoo! est premier en termes de parts de marché sur le "search", notamment grâce au site [Yahoo! Knowledge +](#) (équivalent de [Yahoo! Answers](#) dans ce pays).

Google n'est pas leader partout, loin de là. Et nous avons de nombreux exemples qui nous montrent que de nombreuses parts de marché peuvent encore être prises et certaines situations modifiées. Rien n'est inéluctable...

Vous basez votre stratégie sur la force de la communauté, c'est cela ?

Oui, tout à fait. La notion de "Web 2.0" change beaucoup de choses en donnant de la force au Web. Cette force, c'est celle de la communauté. Si l'on regarde l'histoire du Web et de l'Internet, les sources d'information, tout comme l'information elle-même, ont évolué. Au départ, l'info venait d'ingénieurs et de scientifiques qui communiquaient sur ce support, notamment dans le cadre des groupes de discussion Usenet. Puis ce sont des webmasters qui ont créé des millions de sites, donnant naissance à des outils comme l'annuaire de Yahoo! ou le moteur d'Altavista. Ces sites ont connu une croissance forte, donnant naissance à des dizaines de milliards de documents disponibles en ligne. Cela a donné naissance à des moteurs de recherche de "deuxième génération" comme Google ou Yahoo! Search Technology. Aujourd'hui, une nouvelle source d'information apparaît avec les communautés qui proposent des trillions d'"objets de connaissance" (*knowledge artifacts*), composants granulaires du Web que des outils comme Flickr ou Yahoo! Answers tentent de structurer.

Notre but est de travailler sur ces "objets de connaissance" et d'obtenir une masse critique d'UGC (*User Generated Content*) de haute qualité pour aider les internautes à mieux trouver l'information. Dans les premiers temps du Web, le support *online* ressemblait beaucoup au schéma du *offline* en ce qui concerne l'information, avec un schéma de type "one to many"... Soit une source d'information vers de nombreux visiteurs. Le Web 2.0 balaye tout cela avec un format "many to many", l'internaute prenant le pouvoir, partageant et explorant de nombreux formats de documents (texte, vidéo, audio, etc.). C'est cette évolution que nous voulons prendre en compte dans nos outils avec des sites comme [Flickr](#), [Del.icio.us](#) ou [Yahoo! Answers](#) qui mettent l'internaute au centre du processus de recherche.

Comment cela se traduit-il sur vos sites ?

Prenez par exemple le site [Yahoo! Movies](#) en Grande-Bretagne (voir illustration ci-contre). L'utilisateur y a accès à de nombreuses informations sur ses films favoris :

- Les présentations et bandes annonces "officielles".
- Des infos fournies par la communauté : photos, avis, critiques, votes, etc.
- Des systèmes de questions-réponses issus de Yahoo! Answers.
- L'actualité "chaude" du domaine.
- Et un moteur de recherche sur le Web pour les données granulaires que l'internaute n'aurait pu trouver *via* les possibilités présentés ci-dessus.

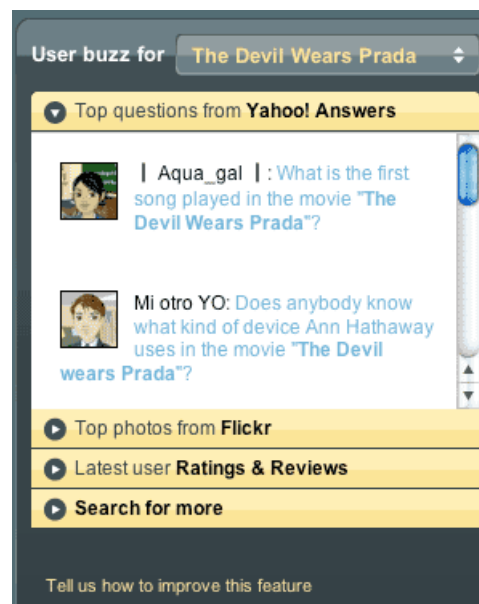
Tous ces outils, toutes ces informations issues de différentes sources, sont complémentaires et sont difficilement dissociables... Notre approche est de les rassembler et de les proposer de la façon la plus efficace possible à l'internaute.

Si je résume en disant que votre stratégie est d'être "aussi bon que Google" pour le search "pur" tout en amenant une "couche" supplémentaire de pertinence grâce à la force de la communauté, est-ce que je me trompe ?

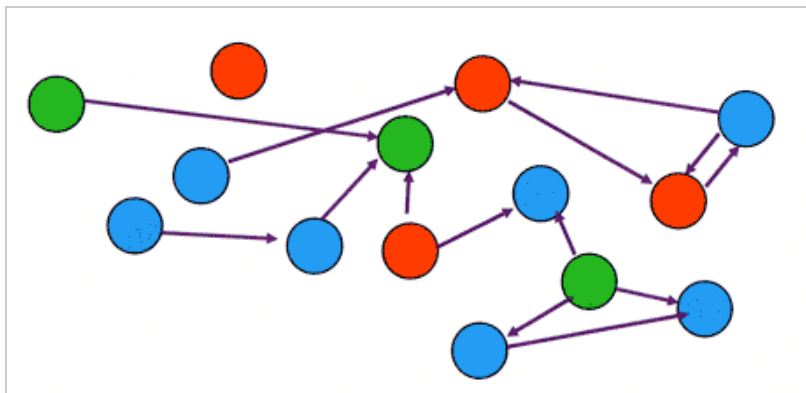
Non, c'est exactement cela. Nous désirons proposer "plus qu'un moteur de recherche". Notre vision se veut plus large que de "simples" algorithmes de pertinence mathématiques. Ceux-ci sont indispensables et primordiaux, certes, mais ils ne suffisent pas. Nous pensons que l'utilisateur doit être mis au centre du processus et pousser de façon bénéfique la qualité des résultats identifiés.

Nous axons nos efforts sur le "social search" et plus particulièrement vers trois directions :

- Le "social search" lui-même avec des outils comme Yahoo! Answers ("Yahoo! Questions/Réponses" en France).
- Le "social bookmarking" grâce à des sites comme Del.icio.us.
- Et le "social media" au travers de partage de fichiers multimedia, grâce à des sites comme Flickr pour la photo.



Par exemple, si vous tapez un mot clé comme "jaguar" sur un moteur "classique", vous aurez de nombreux résultats concernant la marque automobile ou l'animal. Mais les systèmes de "social bookmarking" feront ressortir de façon forte qu'il s'agit aussi du nom d'un des systèmes d'exploitation du Macintosh. Toutes ces sources sont complémentaires et permettent de mettre en avant des résultats qui fournissent une vision globale pour une recherche. Notre idée est que le Web est composé d'"objets" (représentés sur le schéma ci-dessous par des boules de couleur) qui peuvent répondre à une question : pages web, vidéos, chansons, personnes, etc. Ces "objets" sont mis en relation au travers d'hyperliens, de connaissances, de commentaires, d'avis, etc.



Nous essayons de travailler sur des algorithmes qui, sur la base de connaissance de *bons* (en vert) et de *mauvais* (en rouge) objets, doivent apprendre au fur et à mesure la qualité des objets *restants* (en bleu). C'est tout le challenge que nous tentons de relever au quotidien...

N'y a-t-il pas un risque de "spam" dans le social search, avec certains utilisateurs qui diraient certaines choses dans le but de nuire à d'autres ?

Le risque est évident. Mais ce qui fait la force de la communauté, c'est aussi sa capacité à réagir en masse à ce type d'attaque et de tentative de détournement de l'outil. Si nous ajoutons des modérateurs pour tenter également de réagir par nous-même au plus vite, nous pensons que ces dérives peuvent être évitées, dans leur immense majorité tout du moins. Un site comme Flickr pour les photos montre chaque jour que c'est possible !

Quel est le domaine, dans la recherche sur le Web, qui "frémit" le plus actuellement ?

Tout à fait clairement, le monde de la recherche dans les vidéos est celui qui est le plus important. Nous travaillons d'arrache-pied à ce sujet car il représente une vague énorme à court et moyen terme.

Qu'en est-il des annuaires, qui ont fait le succès de Yahoo dans ses premières années ?

Nous restons très attachés aux annuaires de Yahoo! dans le monde entier car ce sont eux qui ont fait en sorte que nous en soyons là aujourd'hui. Cependant, ils ne constituent pas, pour être franc, une priorité à l'heure actuelle pour nos équipes.

Yahoo! a-t-il le plus gros index des moteurs de recherche aujourd'hui ?

Sincèrement, cette question est obsolète. On ne peut plus réellement parler de taille d'index pour quantifier le volume d'informations disponible. Prenons un exemple... On peut estimer qu'un sujet donné est caractérisé en moyenne par 10 pages ou documents (photo, texte, post sur un blog, commentaire, podcast, etc.) sur le Web. On estime qu'une personne vit environ 5 000 "sujets" dans sa vie, ce qui signifie donc qu'une personne est caractérisée par 50 000 "objets" dans une vie. Si on estime à 7 milliards le nombre d'individus sur cette planète, on arrive rapidement à 350 trillions (350 000 000 000 000) d'objets virtuellement accessibles. Un index de 20 milliards de pages représenterait alors 0,0057% de la connaissance humaine. Restons donc modeste et estimons que notre base de données est importante mais pas (encore) exhaustive... Ce comptage peut paraître un peu rapide voire simpliste, certes, mais il montre qu'il est vain de tenter de

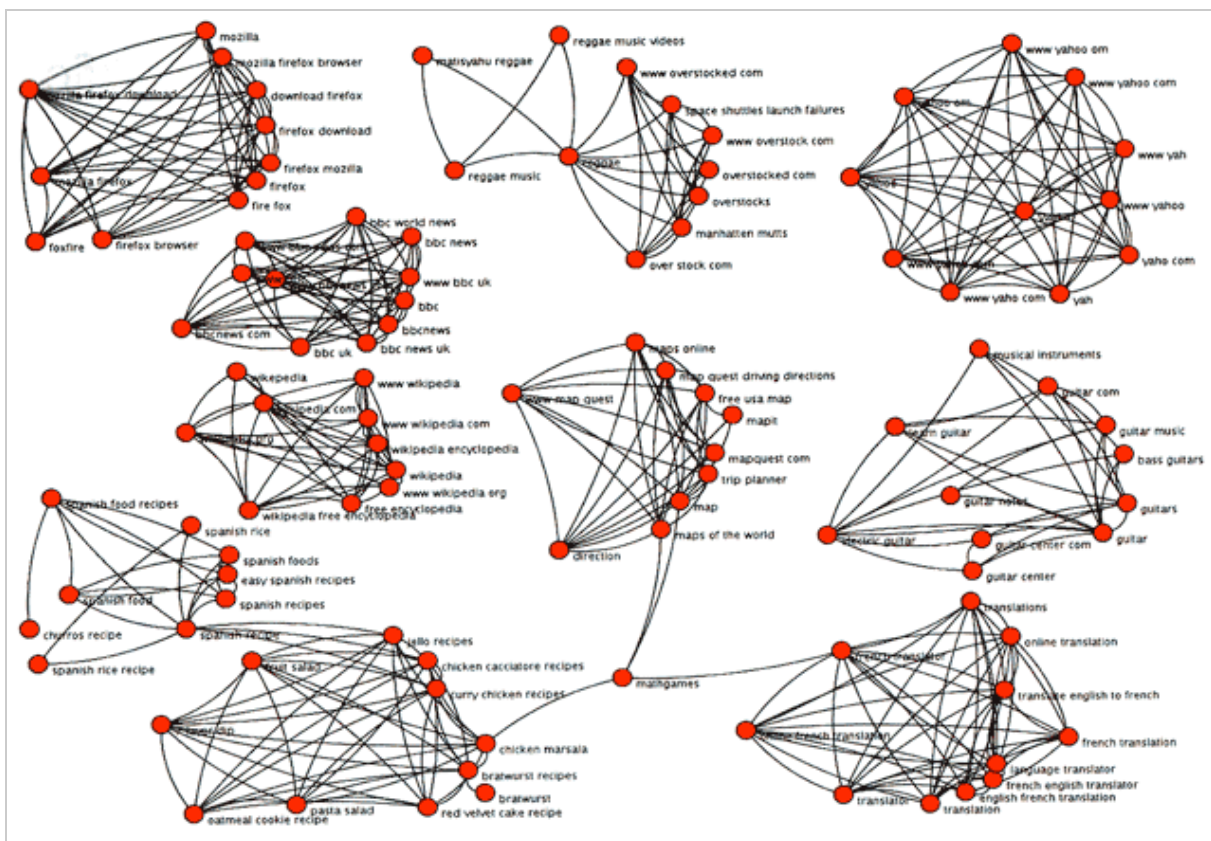
comparer les tailles des index respectifs de chaque acteur, ce que ces derniers ont d'ailleurs arrêté de faire depuis un certain temps... Cela représente peu d'intérêt...

Terminons avec deux questions sur le site de recherche et développement de Yahoo! que vous inaugurez aujourd'hui. Pourquoi l'avoir situé à Barcelone ?

Parce que l'endroit nous semblait central pour attirer de nombreux ingénieurs et spécialistes de toute l'Europe et de l'Amérique du Sud principalement. Et Ricardo Baeza Yates, qui est un expert reconnu du domaine, était sur place. Nous n'avons pas eu d'aide publique spécifique, disons que c'était une question d'opportunité. Douze personnes travaillent d'ores et déjà ici et cette équipe devrait s'étoffer rapidement...

Quel sera le rôle de ce centre ?

Il va travailler sur de la recherche fondamentale dans le cadre de la recherche d'information. L'équipe de Barcelone n'a pas pour charge de travailler sur des produits finis mais sur de réelles tâches de recherche et développement. Voici, par exemple, un schéma d'un de ses travaux actuels sur le "query mining" :



L'idée, ici, est de comprendre puis de grouper en "clusters" de termes la façon dont les internautes recherchent l'information puis de mettre en place des modèles mathématiques permettant de mieux comprendre la façon dont nos visiteurs appréhendent la recherche. C'est un exemple parmi d'autres mais de nombreux travaux sont en cours, comme dans le domaine de la recherche sur terminal mobile, par exemple. Tous ces travaux, toutes ces réflexions, vont aider nos équipes de développement, dont l'une d'entre elles sera basée à Grenoble, en France, dans les prochains temps (et placé sous la responsabilité d'Olivier Parriche, responsable "Search" en France) à élaborer des produits finis qui viendront étoffer les sites de Yahoo!. Le centre de Barcelone est donc, quelque part, garant de ce que l'on trouvera d'ici quelques mois ou années, sur de nombreux sites de Yahoo!...

Merci Stephen.

Merci également aux membre du "staff" de Yahoo! pour leur invitation à Barcelone et leur aide à la réalisation de cet interview.