



L'accès aux contenus multimédia sur Internet

Les moteurs de recherche ont joué un rôle important dans la démocratisation d'Internet et donc dans l'accès à l'information au plus grand nombre. Alors que le nombre de contenus multimédias (photos, vidéos, musique...) explose aujourd'hui sur la toile et dans notre environnement domestique, les outils de recherche ne sont pas encore adaptés. Les moteurs actuels ont en effet été créés pour retrouver des informations textuelles, mais doivent subir des aménagements en profondeur pour devenir également efficaces dans le domaine de l'image (vidéo et photo) et du son. La r&d de France Télécom s'intéresse donc aux moyens les plus pertinents pour effectuer des recherches sur ces nouveaux contenus. De quoi, demain, offrir à ses clients des outils simples et fiables pour avoir accès, où que l'on soit et quel que soit son terminal, à tout type de contenu.

Moteurs de recherche : une évolution indispensable

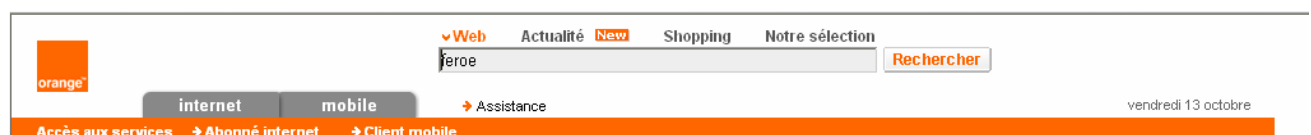
Les effets conjugués de la démultiplication des équipements numériques dans les foyers, du succès du haut débit, des technologies mobiles 3G ont conduit à l'explosion du nombre de contenus multimédias sur Internet, dans les environnements domestiques et sur les portails. Ce phénomène a également été amplifié par l'apparition de nouveaux outils permettant à chacun de s'exposer et de composer sur le Web (blogs, albums photos, contenus autoproduits, etc.), ainsi que par la multiplication de services communautaires.

A leur origine, les moteurs web étaient essentiellement destinés à la recherche de documents textuels et ont rendu populaire le système de recherche par mots clés adossé à une indexation de texte.

Pour pouvoir accompagner le mouvement de consommation croissante de contenus numériques sur différents types de réseaux et de terminaux, et répondre à de nouvelles attentes des internautes, une évolution des moteurs était primordiale, celle-ci est engagée et s'opère en plusieurs étapes progressives.

Une évolution progressive

Une première étape consiste, par déformation du modèle actuel de recherche par mots clés, à utiliser les informations textuelles (descripteurs de contenus, textes d'accompagnement, ...) associées aux contenus multimédias pour les indexer. Cette première évolution incrémentale. Cette première évolution incrémentale répond donc à un souci pragmatique de positionnement stratégique sur ce nouveau flux d'audience en partant de techniques éprouvées.





visionner  vidéos de 1 à 7 sur 7

vendredi 13 octobre 2006

 [Dominique de Villepin en déplacement en Martinique](#)
Dominique de Villepin en déplacement en Martinique - Mediaplayer
1 min 48s - [TF1](#)

lundi 11 septembre 2006

 [Perle Lama, son nouveau clip "Emmène-moi avec toi"](#)
Perle Lama, son nouveau clip "Emmène-moi avec toi" - Musique
3min 39s - [TF1](#)

mercredi 12 juillet 2006

 [Victor Schoelcher - voir la bande-annonce](#)
FRANCE 3 - Victor Schoelcher, un homme contre l'esclavage - Mille et une vie - France 3 Bourgogne Franche-Comté
0min 40s - [France 3](#)

vendredi 07 juillet 2006

lire  dépêché

jeudi 12 octobre 2006

20h59 - Villepin débloque at logement social ..
Dominique de Villepin a ann crédits supplémentaires pot

mercredi 11 octobre 2006

07h41 - Villepin aux Antilles social ..
A six mois de la présidentiel mercredi soir une brève ...

regarder 



La plupart des moteurs ont ainsi d'ores et déjà commencé à intégrer des onglets de recherche spécialisés (verticales) donnant accès à des types de contenus spécifiques (images, vidéo, etc.). Des moteurs spécialisés ont aussi vu le jour en adoptant le même principe d'indexation textuel: [Flickr](#), [Google vidéo](#), etc. France Télécom pour sa part a mis en ligne dans la rubrique Actualités du portail Orange un nouveau moteur de recherche de Journaux Télévisés

Indexer des contenus multimédias

France Télécom s'est engagé à offrir le haut débit sur l'ensemble du territoire. Ceci implique bien sûr de pouvoir proposer des services complets utilisant ces ressources réseau. L'accès simple et pratique aux contenus multimédia, quels que soient les réseaux et les terminaux (simplicité, convergence et "sans couture"), fait donc également partie intégrante de ces objectifs.

Convergence fixe/mobile

L'adaptation des pages html à l'univers des mobiles, tant en termes de format que d'interface, a été longtemps problématique. Les moteurs de recherche ont donc été de ce fait l'apanage du monde des ordinateurs. Avec l'explosion des contenus multimédias, ce problème se pose d'une autre manière : avec des vidéos, des images ou du son, l'adaptation à différentes contraintes de débit et différents formats d'affichage est plus naturelle, puisque souvent préparée en amont. L'accès au haut débit sur les nouvelles générations de terminaux mobiles devenant une réalité, on peut donc parler de véritable convergence fixe-mobile sur les contenus multimédias. La convergence fixe mobile passe aussi par des croisements sur le réseau dans des espaces de stockage de contenus accessibles tant depuis un PC que depuis un terminal mobile.

Octobre 2006



Des méthodes avancées d'analyse des contenus

Alors qu'aujourd'hui, les recherches sur Internet se font essentiellement par mots-clés, le prochain enjeu majeur pour les sites de recherche d'information sera d'améliorer les techniques d'indexation en utilisant non plus seulement les informations textuelles accompagnant le contenu multimedia, mais en analysant le contenu multimedia lui-même. De nombreuses méthodes d'analyse "sémantique" avancée du contenu pourront exister, la r&d de France Télécom y travaille déjà, certaines basées sur l'analyse de la bande son (détection et reconnaissance de locuteurs, transcription audio vers texte, reconnaissance de jingles audio ...), d'autres sur les images ou la vidéo (reconnaissance de plans vidéos, de logos, détection et reconnaissance de visages, reconnaissance des textes dans la vidéo, etc...). Les "méta-données" ainsi produites sont présentées sous forme de texte et permettent de mieux décrire ou retrouver une vidéo, un enregistrement audio ou une photo. Le moteur peut en effet utiliser ces informations pour cibler avec plus de précision et de pertinence ses réponses à une requête.

Même si ces technologies ne sont pas un niveau de maturité tel qu'elles puissent être déployées à des fins d'indexation massive dans les moteurs de recherche elles commencent à apparaître aujourd'hui sur certains d'entre eux.

Au-delà des mots-clés, des requêtes multi-modes

Demain, l'accès aux contenus pourra certainement être possible de manière beaucoup plus naturelle, grâce à des modes de recherche très différents des requêtes par mots-clés couplés à la richesse des interfaces des terminaux nouvelle génération. Les interfaces d'accès aux contenus seront certainement beaucoup plus adhérentes à la nature des contenus eux-mêmes, autorisant une recherche de contenus par l'intermédiaire d'échantillons numériques représentatifs (photo, code barre, ...).

L'audience, un enjeu de taille

Les grands acteurs dans le domaine de la recherche sur Internet ([Yahoo](#), [Google](#), [MSN](#)) drainent une audience extraordinaire, grâce à l'efficacité de leurs moteurs. Après s'être lancés dans une course effrénée pour le nombre de documents indexés (près de 10 milliards de documents pour certains), ceux-ci ont su en effet sensiblement améliorer la pertinence et l'exhaustivité des réponses aux requêtes des internautes. Au-delà de ces énormes volumes de documents indexés, c'est bien sûr l'efficacité des algorithmes de classement des réponses qui fait la différence auprès des internautes. Le ranking mis en œuvre par Google a ainsi contribué à son succès et en a fait une référence sur le marché mondial. Google, surfant sur le succès de son audience ainsi acquise, a par ailleurs su mettre en place un dispositif efficace la rémunérant, jusqu'à en faire une source de revenus essentielle.

Un modèle publicitaire simple et efficace

Dans un premier temps, des mots-clés sont mis aux enchères par une régie et les annonceurs misent une somme sur ceux qui correspondent le mieux à leur activité. Plus le montant proposé est élevé, plus l'annonceur renforce la probabilité, lorsque



Octobre 2006

l'internaute lance une requête comprenant ce mot "acheté", d'apparaître en tête de liste dans la colonne dédiée aux liens publicitaires ou au début de la liste réponse. Lorsque l'internaute clique sur le lien de l'annonceur, celui-ci verse à la régie le montant de la mise sur le mot-clé : c'est le principe du "Coût Par Clic" (CPC). La régie (par exemple [AdWords](#) ou [Yahoo](#) Search Marketing qui sont respectivement les régies possédées en propre par Google et Yahoo) partage ensuite le revenu ainsi généré avec le portail moteur qui a capté l'attention de l'internaute. négocient même leurs services aux portails, selon un partage de revenus sur les mots-clés. Ainsi, il est très fréquent de trouver des "liens sponsorisés" en tête de liste réponse lors d'une requête ou en bannière près de ces listes.

De nouveaux modèles restent à créer

Alors qu'aujourd'hui la monétisation de l'audience sur Internet est un modèle bien établi, il est fort à parier que de nouveaux modèles de rémunération pour les futurs services d'accès aux contenus multimédias seront imaginés. Ainsi, à l'instar des liens sponsorisés sur les moteurs actuels, de nouveaux modèles publicitaires et mode de rémunération de l'audience pourraient générer des revenus. Dans un marché des contenus numériques en pleine croissance, il existe en effet des marges de manœuvre encore importantes pour capter l'audience d'une nouvelle manière. Reste à développer des outils et des services performants pour prendre position rapidement dans ce nouvel espace numérique, le monde d'Internet évoluant très rapidement.

AdWords et AdSense, duo de choc

AdWords, la régie publicitaire du moteur de recherche Google, tire son nom de la contraction des mots "Advertising" (publicité) et "Words" (mots). Grâce à cet outil, des annonces texte ciblées s'affichent lors d'une requête dans le moteur ou sur différents portails (qu'ils aient ou non adopté les technologies moteurs de Google). Cette publicité est très efficace puisqu'elle permet de cibler précisément l'audience en fonction des recherches effectuées sur le moteur. La vente de mots-clés est donc très prisée des annonceurs. Déclinaison Web 2.0 d'AdWords, "AdSense" permet même à tout internaute possédant un site personnel ou un blog d'intégrer un espace publicitaire dans sa page. Les liens proposés sont en rapport direct avec le thème du site. Google paye ensuite à l'internaute un pourcentage du clic payé par l'annonceur. Ces micro-revenus peuvent représenter des sommes colossales si l'on compte les millions de pages perso ou de blogs. C'est le principe de la longue traîne, cher aux théoriciens du Web 2.0.

Un nouveau domaine à explorer

Il apparaît clairement qu'offrir aux clients de France Télécom des services d'accès simples à l'information, qu'elle soit textuelle ou multimédia, est un objectif essentiel pour le Groupe. Alors qu'aujourd'hui les contenus sont de plus en plus hétérogènes



Octobre 2006

et sont disponibles aussi bien sur le Web, dans l'environnement domestique ou sur des portails de recherche spécialisés, proposer des outils simples et efficaces constituerait une valeur ajoutée indéniable. France Télécom est pour cela bien placé, notamment grâce au savoir-faire de son centre de r&d dans les domaines du haut-débit (nécessaire au multimédia) et des mobiles (vecteur puissant du multimédia pour demain). Par ailleurs, sa maîtrise technologique est également incontournable dans les domaines du son, de l'image ou de la vidéo. Les compétences sur l'analyse de l'usage des services, qui peuvent permettre de d'enrichir l'interaction entre les services et l'utilisateur, constituent un autre atout essentiel.

Le Groupe a donc toutes les cartes en main pour pouvoir développer des outils innovants, capables par exemple d'analyser les contenus et de produire de nouveaux descripteurs autorisant des fonctions d'accès d'un nouveau type. Il s'agira de faciliter les modes d'accès à tout type de contenu, quel que soit le terminal utilisé, en s'appuyant sur des outils certainement en rupture par rapport à ce qu'on connaît actuellement sur le marché.

Ainsi, en tirant parti de son savoir-faire technologique, France Télécom pourra capter l'audience de ses clients sur toutes les nouvelles formes d'accès à l'information numérique. La convergence fixe-mobile, réelle dans le domaine des contenus haut-débit, permettra d'autre part de se différencier par rapport à la concurrence en proposant des outils permettant d'accéder, où que l'on soit et quel que soit le terminal, au multimédia. Le Groupe pourra ensuite tirer parti de l'audience ainsi générée grâce à des outils innovants de monétisation.

Liens

[Le dossier du mois de mars 2006 intitulé "Web 2.0 : l'Internet deuxième génération ?"](#)

Glossaire

Ranking : indice de popularité d'une page Internet. Google utilise par exemple son propre "PageRank" (littéralement "rang de classement") pour le classement de ses résultats de recherche. Cette "note" est attribuée en fonction des liens externes (popularité de liens) pointant vers elle et des liens qu'elle fait vers elle-même (liens internes). Dans ce calcul est intégré un algorithme qui a contribué au succès de Google. Le terme "PageRank" fait aussi référence à Larry Page, cofondateur de Google avec Sergey Brin, et qui est l'inventeur de ce principe.

Longue traîne : traduction de "long tail", terme utilisé dans un article de la revue Wired qui a fait référence. Désigne le potentiel économique que représente l'agrégation d'un ensemble de niches. A titre d'exemple, Amazon cite le fait que son chiffre d'affaires sur un grand ensemble d'ouvrages vendus à très peu d'exemplaires est supérieur au chiffre d'affaires qu'il réalise sur les *blockbusters* ou *best-sellers*.